

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM LETRAS
ESTUDOS DA LINGUAGEM
LEXICOGRAFIA, TERMINOLOGIA E TRADUÇÃO: RELAÇÕES TEXTUAIS

Ester Motta

SENTENÇAS JUDICIAIS E LINGUAGEM SIMPLES:

um encontro possível e necessário

Porto Alegre
2022

Ester Motta

SENTENÇAS JUDICIAIS E LINGUAGEM SIMPLES:

um encontro possível e necessário

Texto de tese apresentado como requisito parcial para obtenção do título de Doutora pelo Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul.

Orientadora: Prof.^a Dr.^a Maria José Bocorny Finatto

Porto Alegre
2022

CIP - Catalogação na Publicação

Motta, Ester
Sentenças Judiciais e Linguagem Simples: um
encontro possível e necessário / Ester Motta. --
2022.
411 f.
Orientadora: Maria José Bocorny Finatto.

Tese (Doutorado) -- Universidade Federal do Rio
Grande do Sul, Instituto de Letras, Programa de
Pós-Graduação em Letras, Porto Alegre, BR-RS, 2022.

1. Linguagem Jurídica. 2. Linguagem Simples. 3.
Acessibilidade Textual e Terminológica. 4. Linguística
de Corpus. I. Finatto, Maria José Bocorny, orient.
II. Título.

Ester Motta

SENTENÇAS JUDICIAIS E LINGUAGEM SIMPLES:
um encontro possível e necessário

Texto de tese apresentado como requisito parcial para obtenção do título de Doutora pelo Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul.

Porto Alegre, 13 de junho de 2022.

Resultado: Aprovada.

BANCA EXAMINADORA:

Pof. Dr. Guilherme Fromm
Instituto de Letras e Linguística
Universidade Federal de Uberlândia (UFU)

Profa. Dra. Denise Nauderer Hogetop
Arquivo Público do Estado do Rio Grande do Sul (APERS – SPGG)

Profa. Dra. Aline Evers
Escola de Humanidades
Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS)

APRESENTAÇÃO

Prezado leitor, meu nome é Ester Motta e, neste outono de 2022, aos 59 anos de idade, concluo uma jornada de muita aprendizagem e superação. Lembro que, ao me preparar para o concurso para o cargo de Taquígrafo no Tribunal de Justiça do Estado do Rio Grande do Sul, na década de 80, ouvia de muitas colegas: “a linguagem jurídica é muito difícil”. Como enfrentar desafios e “dar a cara para bater” nunca me detiveram, segui os estudos, prestei o concurso e passei na primeira colocação.

A atividade de taquígrafo, entre outras habilidades, requer o conhecimento da norma padrão da nossa língua portuguesa. E esse conhecimento eu partilhei em muitas aulas que ministrei, particulares ou institucionais em cursos de aperfeiçoamento funcional no TJRS e em outros tribunais.

Sempre gostei de atuar como revisora na Taquigrafia, mas os cursos foram – e ainda são – a minha paixão. Aprendo muito com a atividade de ensino, pois pensar em como tornar determinado conteúdo mais acessível é desafiador. Lecionei para pessoas graduadas e pós-graduadas e pude perceber o quanto muitos textos escritos, pela sua organização sintática ou por seu vocabulário, são complexos e inacessíveis mesmo para esse público.

Essas experiências todas me fizeram buscar, ainda que espaçadamente, o conhecimento. Conciliando as atividades profissionais com as familiares, em 2001, concluí um curso de Especialização no Ensino da Língua Portuguesa na Pontifícia Universidade Católica do Rio Grande do Sul. Em 2010, concluí o Mestrado em Teoria e Análises Linguísticas na Universidade Federal do Rio Grande do Sul. E, agora, em 2022, defendo esta tese de doutorado, que trata de um tema pelo qual me apaixonei desde que com ele me deparei pela primeira vez. E este tema é, grosso modo, a transformação de um texto difícil em um texto fácil.

Espero que este trabalho venha a, junto com tantos outros relacionados ao tema, contribuir para efetivamente diminuir o fosso existente entre a Justiça e o cidadão.

Boa leitura!

AGRADECIMENTOS

Quem anda sozinho pode ir mais rápido, mas nem sempre vai mais longe. Esta frase inicia a música gospel *Cordão de Três Dobras*, que fala sobre a vida a dois. Mas ela pode ser estendida a uma vida a três, a quatro, e a tantos quantos nos deixam marcas positivas, como as de amizade, solidariedade e compreensão.

Apesar de a trajetória de uma pesquisa de doutorado ser um caminho em grande parte solitário, ela não se dá sem o apoio de outras pessoas. Entre essas pessoas, estão os nossos familiares, o nosso orientador, colegas de pesquisa, amigos antigos e amigos que fizemos durante essa caminhada.

Neste momento, então, quero registrar os meus agradecimentos a Jorge Ricardo de Lima Coelho, meu marido e meu incentivador, pela paciência e pelas orientações ao longo da pesquisa. Obrigada por suas lições incansáveis, que me ajudaram a desvendar melhor o complexo mundo conceitual do Direito.

Agradeço ainda:

Às minhas filhas pelo apoio e compreensão. À Mariana, que, mesmo longe, sempre me incentivou a prosseguir com este projeto. E à Vitória, que, estando por perto, sofreu com minha falta de tempo e, por vezes, falta de paciência em razão da sobrecarga de trabalho.

À minha orientadora, Prof^ª. Dra. Maria José Bocorny Finatto, pela paciência, pela confiança e pelo exemplo de profissional que é: sempre na vanguarda nos temas com que trabalha. Muito obrigada por me apresentar a este mundo encantador da Acessibilidade Textual e Terminológica.

A Sidney Leal, pela ajuda com o sistema NILC-Metrix. É um profissional ímpar: além de competente, extremamente simples, humano e solidário. Não foram poucos os *e-mails* que me respondeu sempre prontamente.

A Gabriel Bandle, outro profissional de muita competência e disponibilidade que tive o prazer de conhecer ao longo da pesquisa. Muito obrigada pela presteza em me socorrer com o *Microsoft Office Excel*.

A Elisa Vigna, pelo apoio na revisão final e a formatação do trabalho. Nas palavras de Mário Corso: “Um autor não enxerga a qualidade e o furo de seus escritos”, *Jornal Zero Hora* (10-03-2018).

Aos meus colegas do Grupo de Pesquisa em Acessibilidade Textual e Terminológica, pela parceria. Formam um time animado e sempre na vanguarda como nossa orientadora.

Enfim, agradeço a todos que, direta ou indiretamente, me ajudaram a concluir esta pesquisa.

RESUMO

Esta tese descreve peculiaridades lexicais e sintáticas da linguagem em uso em Sentenças dos Juizados Especiais Cíveis do Poder Judiciário do Estado do Rio Grande do Sul. Tais Juizados se baseiam em princípios como a informalidade e a simplicidade e foram criados para ampliar e facilitar o acesso à Justiça. Nesse cenário, tendo em vista que muitos dos que recorrem a esses Juizados são pessoas com baixa renda e pouca experiência em leitura devido à escolaridade limitada, nossa pesquisa pretende refletir sobre como, pela descrição linguística, se poderia colaborar para promover, por meio dos princípios da Linguagem Simples, a Acessibilidade Textual e Terminológica dessas sentenças. Nossa pesquisa insere-se na Linha de Pesquisa *Lexicografia, Terminologia e Tradução: Relações Textuais* do PPG-Letras-UFRGS e tem como base teórico-metodológica os Estudos da Linguagem e disciplinas afins, tais como: Estudos do Léxico, Linguística de *Corpus*, Linguística Computacional, Linguística Textual, Psicolinguística. Para a consecução das análises, examinamos um conjunto de 440 sentenças desses Juizados Especiais Cíveis em comparação com *corpora* de 4 jornais brasileiros – Folha de São Paulo, Zero Hora, Diário Gaúcho e Jornal Massa –, com um *corpus* de referência do português popular brasileiro escrito, o CorPop, e com um *corpus* de obras literárias dos séculos XIX e XX. As análises se deram de forma manual e também se valeram de ferramentas computacionais. Os resultados sugerem que as sentenças desses Juizados Especiais apresentam padrões lexicais e sintáticos pouco coincidentes com padrões da linguagem cotidiana escrita. Entretanto, com os procedimentos de simplificação textual apresentados no Capítulo 10, constatou-se que se pode tornar esses textos mais acessíveis ao público a que se destinam e assim confirmar que o encontro entre Sentenças Judiciais e Linguagem Simples, além de possível, é extremamente necessário.

Palavras-chave: Linguagem Jurídica; Linguagem Simples; Acessibilidade Textual e Terminológica; Linguística de *Corpus*.

ABSTRACT

This thesis describes lexical and syntactic peculiarities of the language in use in written court decisions of the Special Civil Courts of Rio Grande do Sul. Such Courts are based on principles such as informality and simplicity and were created to expand and facilitate access to Justice. In this scenario, considering that many of those who turn to these courts are people with low income and little experience in reading due to limited education, our research intends to reflect on how, through linguistic description, one could collaborate to promote the textual and terminological accessibility of these written court decisions by using plain language principles. Our research is part of the research line Lexicography, Terminology and Translation: Textual Relations of the PPG-Letras-UFRGS and has as its theoretical and methodological basis the Language Studies and related disciplines, such as: Lexical Studies, *Corpus* Linguistics, Computational Linguistics, Textual Linguistics, Psycholinguistics. To perform the analyses, we examined a set of 440 written court decisions from these Special Civil Courts in comparison with *corpora* from four Brazilian newspapers – Folha de São Paulo, Zero Hora, Diário Gaúcho and Jornal Massa –, with a reference *corpus* of Brazilian popular written Portuguese, CorPop, and with a *corpus* of literary works from the 19th and 20th centuries. The analyses were conducted manually and also employed computational tools. The results suggest that the written court decisions of these Special Courts present lexical and syntactic patterns which have little coincidence with the patterns of written daily language. However, with the text simplification procedures presented in chapter 10, it was verified that it is possible to make these texts more accessible to the intended audience and thus confirm that the encounter between written court decisions and plain language, besides being possible, is extremely necessary.

Keywords: Legal Language; Plain Language; Textual and Terminological Accessibility; *Corpus* Linguistics.

LISTA DE FIGURAS

Figura 1.1 – Exemplo de documento elaborado pela Corregedoria do PJRS antes (E) e depois (D) da Visual Law.....	34
Figura 5.1 – <i>Corpora</i> de Pesquisa	80
Figura 5.2 – Exemplo de Sentença do JEC	83
Figura 5.3 – Zero Hora.....	85
Figura 5.4 – Folha de São Paulo.....	85
Figura 5.5 – Diário Gaúcho.....	86
Figura 5.6 – Jornal Massa	86
Figura 5.7 – Organograma do Poder Judiciário no Brasil.....	93
Figura 5.8 – Passo 1	99
Figura 5.9 – Passo 2	100
Figura 5.10 – Passo 3	100
Figura 5.11 – Passo 4	101
Figura 5.12 – Passo 5	101
Figura 5.13 – Passo 6	101
Figura 5.14 – Passo 7	102
Figura 5.15 – Passo 8.....	102
Figura 5.16 – Sistema SIMPLIFICA, do Projeto PorSimples do NILC-USP.....	111
Figura 5.17 – Tela Inicial do NILC-Metrix.....	113
Figura 5.18 – Espaço para colocar o texto a ser analisado pelo NILC-Metrix	113
Figura 5.19 – Tela com os resultados da análise pelo NILC-Metrix.....	114
Figura 9.1 – Exemplo de <i>Types</i> do CE e do Jornal ZH.....	285
Figura 9.2 – Percentuais de <i>Types</i> Exclusivos e Duplicados do CE e do Jornal ZH.....	286
Figura 9.3 – Percentuais de <i>Types</i> Exclusivos e Duplicados do CE e do Jornal FSP.....	291
Figura 9.4 - Percentuais de <i>Types</i> Exclusivos e Duplicados do CE e do JM	297
Figura 9.5 – Percentuais de <i>Types</i> Exclusivos e Duplicados do CE e do DG.....	300
Figura 9.6 – Percentuais de <i>Types</i> Exclusivos e Duplicados do CE e do CorPop.....	310
Figura 9.7 – Percentuais de <i>Types</i> Exclusivos e Duplicados do CE e do CR1	315
Figura 10.1 – Capa e contracapa do Guia de Linguagem Simples do TJRS.....	368
Figura 10.2 – Página 2 do Guia de Linguagem Simples do TJRS	368
Figura 11.1– Imagem de uma sentença da Justiça do Trabalho em <i>Visual Law</i>	387
Figura A.1– Passo 1	409

Figura A.2 – Passo 2	409
Figura A.3 – Passo 3	410

LISTA DE GRÁFICOS

Gráfico 6.1 – Índice Flesch	121
Gráfico 6.2 – Cláusula por Sentença.....	123
Gráfico 6.3 – Sentenças por Parágrafo x Palavras por Sentença.....	124
Gráfico 6.4 – Sílabas por Palavra de Conteúdo	125
Gráfico 6.5 – Palavras antes dos Verbos Principais	126
Gráfico 6.6 – Incidência de Negações	127
Gráfico 6.7 – Incidência de Conectivos	128
Gráfico 6.8 – Sobreposição de Argumentos Adjacentes.....	128
Gráfico 6.9 – Referência Anafórica Adjacente	130
Gráfico 6.10 – Frequências de Palavras de Conteúdo	131
Gráfico 6.11 – Índice Flesch x Demais Métricas	132
Gráfico 7.1 – Tipos de CFSs em Percentuais	183
Gráfico 8.1 – Densidade Semântica x Palavras Concretas	242
Gráfico 8.2 – Densidade Semântica x Idade de Aquisição x Palavras Concretas.....	244
Gráfico 8.3 – Orações Relativas x Pronomes Relativos x Pronomes Totais.....	257
Gráfico 8.4 – Incidência de Conectivos Total x Proporção de Conjunções Subordinativas em relação a todas as Conjunções do Texto.	262
Gráfico 8.5 – Proporção de sujeitos pospostos em relação a todos os sujeitos do texto e Proporção de orações que não estão no formato SVO (sujeito-verbo-objeto) em relação a todas as orações do texto	265
Gráfico 8.6 – Referentes x Pronomes	271
Gráfico 8.7 – Resultado dos Grupos de Métricas nos <i>Corpora</i>	278
Gráfico 9.1 – Proporção de <i>Types</i> Exclusivos entre o CE e o Jornal ZH	290
Gráfico 9.2 – Proporção de <i>Types</i> Exclusivos entre CE e o Jornal FSP.....	294
Gráfico 9.3 – Proporção de <i>Types</i> Exclusivos entre o CE e o JM.....	299
Gráfico 9.4 – Proporção de <i>Types</i> Exclusivos do CE e do DG	303
Gráfico 9.5 – Percentual de <i>Types</i> Exclusivos no CE e nos Jornais Analisados.....	304
Gráfico 9.6 – Percentual de <i>Types</i> Exclusivos do CE.....	305
Gráfico 9.7 – Percentual de <i>Types</i> Exclusivos entre o CE e o CorPop.....	312
Gráfico 9.8 – Percentual de <i>Types</i> Exclusivos do CE em relação aos <i>Corpora</i> de Contraste	314
Gráfico 9.9 – Percentual de <i>Types</i> Exclusivos do CE.....	318

LISTA DE QUADROS

Quadro 4.1 – Caracterização dos Tipos de <i>Corpora</i>	58
Quadro 4.2 – Termos empregados com frequência neste trabalho.....	61
Quadro 5.1 – Caracterização do CorPop (CCP)	88
Quadro 5.2 – Conteúdo do <i>Corpus</i> de Referência 1 (CR1)	89
Quadro 5.3 – Caracterização do CR1	90
Quadro 5.4 – Caracterização do CR2	91
Quadro 5.5 – Caracterização do CE.....	98
Quadro 5.6 – Caracterização dos <i>Corpora</i> de Contraste.....	106
Quadro 6.1 – Exemplos de Lemas das Sentenças e do DG Inexistentes no CorPop.....	138
Quadro 6.2 – Reescrita 1 - Incolumidade.....	140
Quadro 6.3 – Reescrita 2 - Incolumidade.....	141
Quadro 6.4 – Reescrita 3 - Incolumidade.....	142
Quadro 6.5 – Reescrita 4 - Incolumidade.....	142
Quadro 6.6 – Reescrita 5 - Incolumidade.....	143
Quadro 6.7 – Escores do Trecho Original e Reescrita 5 - Incolumidade.....	143
Quadro 6.8 – Reescrita 1 - Sinistralidade.....	144
Quadro 6.9 – Reescrita 2 - Sinistralidade.....	145
Quadro 6.10 – Reescrita 3 - Sinistralidade.....	146
Quadro 6.11 – Escores do Trecho Original e Reescrita 3 - Sinistralidade.....	147
Quadro 7.1 – Problemas de Análise no Parser Palavras	174
Quadro 7.2 – Exemplo de frase complexa	176
Quadro 7.3 – Exemplos de CFSs Pós-Verbais das Sentenças da Amostra 2	209
Quadro 7.4 – Exemplos de CFSs Pré-Verbais das Amostras.....	212
Quadro 7.5 – Exemplos de CFSs Pós-Verbais das Amostras	213
Quadro 7.6 – Exemplos de CFSs Oracionais das Amostras	215
Quadro 7.7 – Exemplos de CFSs Elípticos das Amostras	217
Quadro 7.8 – Exemplos de Tipos de CFSs Oracionais das Amostras	218
Quadro 7.9 – Exemplos de CFSs Pronominais das Amostras.....	220
Quadro 7.10 – Exemplos de CFCVs das Amostras.....	223
Quadro 7.11 – Exemplos de CFCVs das Amostras com mais de 30 Palavras.....	226
Quadro 7.12 – Exemplos de CFCVs Oracionais das Amostras	229
Quadro 8.1 – Exemplos do CE e do DG.....	247

Quadro 8.2 – Exemplos do CE	251
Quadro 8.3 – Exemplos do DG	253
Quadro 9.1 – Lista de Palavras Duplicadas entre os 30 primeiros <i>types</i> do CE e do Jornal ZH	283
Quadro 9.2 – Lista de Palavras Exclusivas do CE e do Jornal ZH entre os 30 primeiros <i>types</i>	284
Quadro 9.3 – Exemplos de Palavras Duplicadas no CE e no Jornal ZH	288
Quadro 9.4 – Sugestão de Simplificação de Modo Individualizado.....	308
Quadro 9.5 – Acepções do Dicionário Houaiss para o verbete <i>demandar</i>	327
Quadro 9.6 – Acepções do Dicionário Houaiss para o verbete <i>demanda</i>	327
Quadro 9.7 – Acepções do Dicionário Houaiss para os verbetes <i>demandada</i> e <i>demandante</i>	327
Quadro 9.8 – Acepções do Dicionário Houaiss para o verbete <i>homologar</i>	328
Quadro 10.1 – Acepções do Dicionário Houaiss para os verbetes <i>improcedência</i> e <i>improcedente</i>	340
Quadro 10.2 – Reescritas Sugeridas para o Trecho 1	341
Quadro 10.3 – Exemplos de palavras a evitar e substituir	364

LISTA DE TABELAS

Tabela 4.1 – Níveis de Leiturabilidade segundo a Fórmula do Índice Flesch	66
Tabela 5.1 – Conteúdo do CorPop.....	88
Tabela 5.2 – Sentenças do CE por Período	96
Tabela 5.3 – Número de palavras do CE por Turma Recursal e por Período	97
Tabela 5.4 – Sentenças do CE por Temática.....	98
Tabela 5.5 – Número de <i>Tokens</i> e <i>Types</i> por Autor do CR1.....	107
Tabela 5.6 – Número de <i>Tokens</i> e <i>Types</i> do CE, dos <i>Corpora</i> de Contraste e de Referência	109
Tabela 6.1 – Quantidade de Palavras dos <i>Corpora</i> no Estudo-Piloto 1.....	118
Tabela 6.2 – Índices das Métricas do Coh-Metrix <i>Dementia</i> para o Estudo-Piloto 1	119
Tabela 6.3 – Índices Sentenças por Parágrafo e Palavras por Sentença	124
Tabela 6.4 – Hierarquia de Legibilidade dos <i>Corpora</i>	131
Tabela 6.5 – Comparação das Médias da Tabela 6.4 com o Índice Flesch.....	132
Tabela 6.6 – Dados dos <i>corpora</i> extraídos do AntConc e do Coh-Metrix-Port.....	136
Tabela 6.7 – Comparação entre os Lemas dos <i>Corpora</i>	137
Tabela 6.8 – Respondentes por Faixa Etária	156
Tabela 6.9 – Respondentes por Escolaridade	156
Tabela 6.10 – Respondentes por Renda Familiar	158
Tabela 6.11 – Escore da Questão 4 - Aduziu	160
Tabela 6.12 – Escore da Questão 5 - Distinção.....	160
Tabela 6.13 – Escore da Questão 6 - Readequação	160
Tabela 6.14 – Escore da Questão 7 - Sinistralidade	161
Tabela 6.15 – Escore da Questão 8 - Faixa Etária.....	161
Tabela 6.16 – Respostas Corretas das Questões 4 a 8 em Ordem Decrescente	161
Tabela 6.17 – Respostas Dadas à Questão 9	163
Tabela 6.18 – Respostas Dadas à Questão 10 – Escrita.....	164
Tabela 6.19 – Respostas Dadas à Questão 11 – Escrita 2.....	164
Tabela 6.20 – Respostas Dadas à Questão 12 – Escrita 3.....	165
Tabela 6.21 – Respostas das Questões 10, 11 e 12 por Ordem de Dificuldade	165
Tabela 7.1 – Total de Sentenças por Temática e Quantidade de Sentenças para a Amostra 1	170
Tabela 7.2 – Total de Sentenças por Turma Recursal	170
Tabela 7.3 – Total de Sentenças por Ano de Coleta.....	171

Tabela 7.4 – Dados do CE e da Amostra 1	171
Tabela 7.5 – Reescritas Simplificadas da Frase Constante no Quadro 7.2.....	177
Tabela 7.6 – Ocorrências por Tipos de CFCVs da Amostra	182
Tabela 7.7 – CFSs Pré-Verbais por Número de Palavras	184
Tabela 7.8 – CFSs Pós-Verbais por Número de Palavras	186
Tabela 7.9 – CFSs Oracionais por Número de Palavras	188
Tabela 7.10 – Tipos de CFSs Oracionais	189
Tabela 7.11 – Tipos de CFSs Pronominais	190
Tabela 7.12 – Formas Verbais Referentes aos CFSs Elípticos	192
Tabela 7.13 – Constituintes Frasais com Função de Complemento Verbal (CFCVs)	195
Tabela 7.14 – CVFCVs e seus Subtipos Presentes na Amostra 1	196
Tabela 7.15 – Tipos de CFCVs Oracionais.....	197
Tabela 7.16 – Ocorrência de CFCVs por Tipo e Número de Palavras.....	199
Tabela 7.17 – Dados das Amostras 1 e 2 do CE.....	204
Tabela 7.18 – Total de Sentenças por Temática nas Amostras 1 e 2.....	204
Tabela 7.19 – Dados das Amostras 2, do DG e do JM	205
Tabela 7.20 – Ocorrências por Tipos de CFSs das Amostra 1, DG e JM.....	206
Tabela 7.21 – CFSs por Número de Palavras na Amostra 2, DG e JM	211
Tabela 7.22 – Formas Verbais Referentes aos CFSs Elípticos	216
Tabela 7.23 – Tipos de CFSs Oracionais	217
Tabela 7.24 – Tipos de CFSs Pronominais	219
Tabela 7.25 – Constituintes Frasais com Função de Complemento Verbal (CFCVs)	221
Tabela 7.26 – CVFCVs e seus Subtipos Presentes nas Amostra 2, DG e JM.....	222
Tabela 7.27 – CFCVs Sintagmas Nominais por número de palavras.....	224
Tabela 7.28 – Percentuais de CFCVs Sintagmas Nominais por Número de Palavras	225
Tabela 7.29 – CFCVs Oracionais por Número de Palavras	227
Tabela 7.30 – Percentuais de CFCVs Oracionais por Número de Palavras.....	227
Tabela 7.31 – Tipos de CFCVs Oracionais.....	228
Tabela 8.1 – Dados do CE e dos <i>Corpora</i> de Contraste	236
Tabela 8.2 – Densidade Semântica	239
Tabela 8.3 – Média de Palavras Concretas no Texto.....	241
Tabela 8.4 – Média da Idade de Aquisição das Palavras do Texto	243
Tabela 8.5 – Proporção de Adjetivos em relação às demais Palavras do Texto	245
Tabela 8.6 – Índice Flesch do CE e <i>Corpora</i> de Contraste.....	246

Tabela 8.7 – Distância na Árvore de Dependência.....	248
Tabela 8.8 – Número de Palavras antes dos Verbos Principais.....	250
Tabela 8.9 – Incidência de Pronomes	252
Tabela 8.10 – Proporção de Orações Relativas em relação à Quantidade de Orações do Texto	254
Tabela 8.11 – Comparação entre a Proporção de Orações Relativas em relação à Quantidade de Orações do Texto, a Incidência de Pronomes em geral e a Incidência de Pronomes Relativos	256
Tabela 8.12 – Incidência de Conectivos Total	258
Tabela 8.13 – Proporção de Orações Subordinadas pela Quantidade de Orações do Texto e Proporção de Conjunções Subordinativas em relação a todas as Conjunções do Texto	260
Tabela 8.14 – Incidência de Conectivos Total x Proporção de Conjunções Subordinativas em relação a todas as Conjunções do Texto	262
Tabela 8.15 – Proporção de Sujeitos Pospostos em relação a todos os Sujeitos do Texto e Proporção de Orações que não estão no formato SVO (sujeito-verbo- objeto) em relação a todas as Orações do Texto.....	264
Tabela 8.16 – Quantidade Média de Referentes que se repetem nos Pares de Sentenças Adjacentes do Texto	267
Tabela 8.17 – Quantidade Média de Referentes que se repetem nos Pares de Sentenças do Texto	269
Tabela 8.18 – Proporção de Palavras que Denotam Negação em relação à Quantidade de Palavras do Texto.....	272
Tabela 8.19 – Média das Métricas Analisadas	273
Tabela 8.20 – Média das Métricas Lexicais	275
Tabela 8.21 – Média das Métricas de Leitabilidade.....	275
Tabela 8.22 – Média das Métricas Sintáticas	277
Tabela 9.1 – Número de <i>Tokens</i> e <i>Types</i> do CE, dos <i>Corpora</i> de Contraste e de Referência	280
Tabela 9.2 – Exemplo de Comparação entre Listas de Palavras	281
Tabela 9.3 – Dados dos <i>Types</i> do CE e do Jornal ZH.....	286
Tabela 9.4 – Distribuição dos 20 <i>Types</i> mais frequentes na Comparação entre o CE e o Jornal ZH	286
Tabela 9.5 – <i>Types</i> Exclusivos do CE e do Jornal ZH.....	289
Tabela 9.6 – Dados dos <i>Types</i> do CE e do Jornal FSP	291

Tabela 9.7 – Distribuição dos 20 <i>types</i> mais frequentes na comparação entre o CE e o Jornal FSP	292
Tabela 9.8 – <i>Types</i> Exclusivos do CE e do Jornal FSP	294
Tabela 9.9 – Dados dos <i>Types</i> do CE e do JM	296
Tabela 9.10 – Distribuição dos 20 <i>Types</i> mais Frequentes na Comparação entre o CE e o JM	297
Tabela 9.11 – <i>Types</i> Exclusivos do CE e do JM	299
Tabela 9.12 – Dados dos <i>Types</i> do CE e do DG.....	300
Tabela 9.13 – Distribuição dos 20 <i>Types</i> mais Frequentes na Comparação entre o CE e o DG	301
Tabela 9.14 – <i>Types</i> Exclusivos do CE e do DG.....	303
Tabela 9.15 – Percentual de <i>Types</i> Exclusivos do CE.....	305
Tabela 9.16 – Dados dos <i>Types</i> do CE e do CorPop	310
Tabela 9.17 – Distribuição dos 20 <i>Types</i> mais frequentes na Comparação entre o CE e o CorPop	311
Tabela 9.18 – <i>Types</i> Exclusivos do CE e do CorPop.....	312
Tabela 9.19 – Percentual de <i>Types</i> Exclusivos do CE em relação aos <i>Corpora</i> de Contraste	313
Tabela 9.20 – Dados dos <i>Types</i> do CE e do CR1	315
Tabela 9.21 – Distribuição dos 20 <i>Types</i> mais frequentes na Comparação entre o CE e o CR1	315
Tabela 9.22 – <i>Types</i> Exclusivos do CE.....	317
Tabela 9.23 – Percentuais de <i>Types</i> Exclusivos do CE.....	318
Tabela 9.24 – Vinte Primeiras Palavras-Chave do CE x CR1	320
Tabela 9.25 – Frequência dos verbos <i>opinar, homologar, produzir e considerar</i> nos <i>Corpora</i> de Contraste e no CR1	322
Tabela 9.26 – Vinte Primeiras Palavras-Chave do CE x CR2	324
Tabela 9.27 – Frequência dos verbos <i>intimar e publicar</i> nos <i>Corpora</i> de Contraste e nos <i>Corpora</i> de Referência.....	325
Tabela 9.28 – <i>Types</i> Exclusivos do CE em relação a todos os <i>corpora</i> analisados	326
Tabela 9.29 – Frequência de <i>Types</i> da Família de <i>demandada/demandante</i> e <i>homologar/homologatória</i> nos <i>Corpora</i> de Contraste e nos <i>Corpora</i> de Referência	329
Tabela 9.30 – Vinte Primeiras Multipalavras-Chave do CE x CR1	330
Tabela 9.31– Vinte Primeiras Multipalavras-Chave do CE x CR2.....	332

Tabela 9.32 – Vinte Primeiras Multipalavras mais Frequentes do CE Duplicadas nas Listas de Multipalavras do CR1 e do CR2.....	333
Tabela 10.1 – Frequência dos <i>Types</i> sublinhados no Trecho 2.....	344
Tabela 10.2 – Índices do Coh-Metrix na Sentença Original e na Reescrita.....	359

LISTA DE ABREVIATURAS E SIGLAS

ACD – Análise Crítica do Discurso

AMB – Associação dos Magistrados Brasileiros

ASL – Análise Semântica Latente

AT – Acessibilidade Textual

ATT – Acessibilidade Textual e Terminológica

CCA – *Corpus* de Contraste A | *corpus* de jornais tradicionais – Zero Hora e Folha de São Paulo

CCB – *Corpus* de Contraste B | *corpus* de jornais populares – Diário Gaúcho e Massa

CCP – *Corpus* de Contraste | CorPop composto por textos jornalísticos e literários com base no nível de letramento médio dos leitores do país e representativos do Português Popular Escrito

CDC – Código de Defesa do Consumidor

CE – *Corpus* de Estudo

CFSs – Constituintes frasais com função de sujeito

CFCVs – Constituintes frasais com função de complementos verbais

CFCVPs – Complementos com função de complementos verbais preposicionados

CFCVNPs – Complementos com função de complementos verbais não preposicionados

CFCVPRs – Complementos com função de complementos verbais representados por pronomes

CLEF – Cross- Language Education and Function

CMLJ– Comissão de Modernização da Linguagem Jurídica

CR1 – *Corpus* de Referência 1 | obras literárias brasileiras que estão em domínio público

CR2 – *Corpus* ptTenTen: *Corpus of the Portuguese Web*

CT – Complexidade textual

DG – Diário Gaúcho

DS – Densidade Semântica

EJA – Educação de Jovens e Adultos

FGV – Fundação Getúlio Vargas

FSP – Folha de São Paulo

IF – Índice Flesch

INAF – Indicador de Alfabetismo Funcional

IPESPE – Instituto de Pesquisas Sociais, Políticas e Econômicas

JECs – Sentenças dos Juizados Especiais Cíveis
JFRN – Justiça Federal do Rio Grande do Norte
JM – Jornal Massa
LA – Linguística Aplicada
LC – Linguística de *Corpus*
LCOMP – Linguística Computacional
LINDB – Lei de Introdução às Normas do Direito Brasileiro
LS – Linguagem Simples
MS – Ministério da Saúde
NILC – Núcleo Interinstitucional de Linguística Computacional
PJRS – Poder Judiciário do Estado do Rio Grande do Sul
PLN – Processamento da Linguagem Natural
PPE – Português Popular Escrito
REBRALS – Rede Brasileira de Letramento em Saúde
SAPro – Sujeitos Anafóricos Pronominais
SN – Sintagma nominal
SP1 – Sentenças-Piloto 1
SP2 – Sentenças-Piloto 2
ST – Simplificação Textual
SV – Sintagma verbal
SVO – Sujeito-verbo-objeto
TCC – Trabalho de conclusão de curso de graduação
TCT – Teoria Comunicativa da Terminologia
TJDFT – Tribunal de Justiça do Distrito Federal e dos Territórios
TTR – Relação *types/tokens*
USP – Universidade de São Paulo
ZH – Jornal Zero Hora
1TRC – Primeira Turma Recursal Cível
2TRC – Segunda Turma Recursal Cível
3TRC – Terceira Turma Recursal Cível
4TRC – Quarta Turma Recursal Cível

SUMÁRIO

1	INTRODUÇÃO	26
1.1	ESTA PESQUISA E OS MOVIMENTOS EM FAVOR DA SIMPLICIDADE NA LINGUAGEM JURÍDICA NO BRASIL	29
1.2	ESTA PESQUISA E OS ESTUDOS LINGÜÍSTICOS NO BRASIL	38
2	TEMA DA PESQUISA.....	43
2.1	QUESTÕES DE PESQUISA.....	44
2.2	HIPÓTESES DE PESQUISA	44
3	OBJETIVOS.....	47
4	PRESSUPOSTOS TEÓRICOS.....	49
4.1	A TERMINOLOGIA.....	49
4.1.1	Terminologia Textual	51
4.2	A LINGÜÍSTICA DE <i>CORPUS</i>	53
4.3	A ACESSIBILIDADE TEXTUAL	62
4.3.1	Conceitos relacionados à Acessibilidade Textual	63
4.3.2	Estudos referentes à Acessibilidade Textual e aos conceitos a ela relacionados ..	64
4.3.3	A Acessibilidade Textual e Terminológica.....	69
4.3.4	A Simplificação Textual e o Texto Especializado	70
4.4	A TRADUÇÃO INTRALINGÜÍSTICA E A SIMPLIFICAÇÃO TEXTUAL.....	72
4.5	ESTUDOS RELACIONADOS A ESTA PESQUISA	74
5	MATERIAIS E MÉTODOS	79
5.1	DESCRIÇÃO DOS <i>CORPORA</i>	81
5.1.1	<i>Corpus</i> de Estudo - CE.....	81
5.1.2	<i>Corpus</i> de Contraste CCA	84
5.1.3	<i>Corpus</i> de Contraste CCB	85
5.1.4	<i>Corpus</i> de Contraste CCP.....	87
5.1.5	<i>Corpus</i> de Referência 1 – CR1.....	89
5.1.6	<i>Corpus</i> de Referência 2 – CR2.....	91
5.2	COMPILAÇÃO DOS <i>CORPORA</i> CE, CCA, CCB E CR1.....	92
5.2.1	<i>Corpus</i> de Estudo	92
5.2.2	<i>Corpus</i> de Contraste CCA	103
5.2.3	<i>Corpus</i> de Contraste CCB	105

5.2.4	<i>Corpus de Referência 1 – CR1</i>	107
5.3	FERRAMENTAS UTILIZADAS NAS ANÁLISES.....	109
6	ESTUDOS PARCIAIS 1	116
6.1	ESTUDO-PILOTO 1 – PRIMEIRA EXPERIÊNCIA COM AS MÉTRICAS DA FERRAMENTA COH-METRIX-PORT	116
6.1.1	Constituição dos <i>corpora</i> do Estudo-Piloto 1	116
6.1.2	Processamento e Análise dos Dados	118
6.2	ESTUDO-PILOTO 2 – ÍNDICES LEXICAIS E PROPOSTAS DE SIMPLIFICAÇÃO DE TRECHOS DE UMA AMOSTRA DE SENTENÇAS X UMA AMOSTRA DO DG X CORPOP	134
6.2.1	Constituição dos <i>corpora</i> do Estudo-Piloto 2	135
6.2.2	Processamento e Análise dos Dados	135
6.2.3	Propostas de Reescrita e/ou Tradução Intralinguística	139
6.3	QUESTIONÁRIO DE VERIFICAÇÃO DA COMPLEXIDADE TEXTUAL.....	148
6.3.1	Análise dos Dados	155
7	ESTUDOS PARCIAIS 2	169
7.1	ESTUDO-PILOTO 3 – CONSTITUINTES FRASAS DA AMOSTRA 1 (110 SENTENÇAS DO CE)	169
7.1.1	Constituição da amostra do CE	169
7.1.2	Processamento e Análise dos Dados	173
7.2	ESTUDO-PILOTO 4 – CONSTITUINTES FRASAS EM AMOSTRAS DO CE, DO DG E DO JM	203
7.2.1	Constituição das amostras	203
7.2.2	Processamento e Análise dos Dados	206
8	MÉTRICAS DO NILC-METRIX ANALISADAS	235
8.1	DENSIDADE SEMÂNTICA.....	238
8.2	MÉDIA DE PALAVRAS CONCRETAS NO TEXTO	240
8.3	(#) MÉDIA DA IDADE DE AQUISIÇÃO DAS PALAVRAS DO TEXTO	243
8.4	INCIDÊNCIA DE ADJETIVOS.....	244
8.5	ÍNDICE FLESCH.....	246
8.6	DISTÂNCIA NA ÁRVORE DE DEPENDÊNCIAS.....	247
8.7	NÚMERO DE PALAVRAS ANTES DE VERBOS PRINCIPAIS.	249
8.8	INCIDÊNCIA DE PRONOMES.....	252

8.9	(#) PROPORÇÃO DE ORAÇÕES RELATIVAS EM RELAÇÃO À QUANTIDADE DE ORAÇÕES DO TEXTO.....	254
8.10	INCIDÊNCIA DE CONECTIVOS TOTAL.....	258
8.11	(#) PROPORÇÃO DE ORAÇÕES SUBORDINADAS PELA QUANTIDADE DE ORAÇÕES DO TEXTO E PROPORÇÃO DE CONJUNÇÕES SUBORDINATIVAS EM RELAÇÃO A TODAS AS CONJUNÇÕES DO TEXTO.....	260
8.12	(#) PROPORÇÃO DE SUJEITOS POSPOSTOS EM RELAÇÃO A TODOS OS SUJEITOS DO TEXTO.....	263
8.13	(#) PROPORÇÃO DE ORAÇÕES QUE NÃO ESTÃO NO FORMATO SVO (SUJEITO-VERBO OBJETO) EM RELAÇÃO A TODAS AS ORAÇÕES DO TEXTO...	264
8.14	QUANTIDADE MÉDIA DE REFERENTES QUE SE REPETEM NOS PARES DE SENTENÇAS ADJACENTES DO TEXTO.....	266
8.15	QUANTIDADE MÉDIA DE REFERENTES QUE SE REPETEM NOS PARES DE SENTENÇAS DO TEXTO.....	268
8.16	PROPORÇÃO DE PALAVRAS QUE DENOTAM NEGAÇÃO EM RELAÇÃO À QUANTIDADE DE PALAVRAS DO TEXTO	272
9	PADRÕES LEXICAIS DOS <i>CORPORA</i> ANALISADOS.....	280
9.1	CE X CCA	285
9.1.1	CE X ZH	285
9.1.2	CE X FSP	291
9.2	CE X CCB.....	296
9.2.1	CE X JM	296
9.2.2	CE X DG	300
9.3	CE X CORPOP	309
9.4	CE X <i>CORPORA</i> DE REFERÊNCIA (CR1 E CR2).....	314
9.4.1	CE X CR1 – PALAVRAS-CHAVE	319
9.4.2	CE X CR2 – PALAVRAS-CHAVE	323
9.4.3	CE X CR1 – MULTIPALAVRAS.....	330
9.4.4	CE X CR2 – MULTIPALAVRAS.....	332
10	PROPOSTAS DE SIMPLIFICAÇÃO TEXTUAL	338
10.1	PROPOSTAS DE REESCRITA	338
10.1.1	Improcedência	339
10.1.2	<i>Quantum</i>.....	343
10.1.3	Exemplar de Sentença	346

10.2	ORIENTAÇÕES BÁSICAS PARA ESCREVER EM LINGUAGEM SIMPLES ...	360
10.2.1	Empatia – Coloque-se no lugar do leitor	361
10.2.2	Organize o texto.....	361
10.2.3	Evite excesso de remissões	362
10.2.4	Evite que seu leitor tenha de procurar informações fora do texto.....	363
10.2.5	Use palavras simples (mais familiares e menos complexas)	363
10.2.6	Evite palavras desnecessárias.....	364
10.2.7	Use mais verbos	364
10.2.8	Use frases mais curtas.....	365
10.2.9	Coloque, sempre que possível, a frase em ordem direta (sujeito-verbo-objeto).	366
11	CONCLUSÕES E PERSPECTIVAS.....	370
11.1	RETOMADA DAS QUESTÕES DE PESQUISA	376
11.2	RETOMADA DAS HIPÓTESES DE PESQUISA.....	383
11.3	PERSPECTIVAS	386
	REFERÊNCIAS.....	392
	ANEXOS	404
	ANEXO 1 - TEXTO ORIGINAL DA SENTENÇA SIMPLIFICADA NO CAPÍTULO 10	404
	ANEXO 2 - TEXTO SIMPLIFICADO DO EXEMPLAR DA SENTENÇA DO CAPÍTULO 10	407
	ANEXO 3 – <i>CORPUS</i> DE ESTUDO	409
	ANEXO 4 - LISTAS DE PALAVRAS UTILIZADAS NA PESQUISA.....	411
	<i>CORPUS</i> DE ESTUDO	411
	<i>CORPUS</i> DE CONTRASTE DO JORNAL DIÁRIO GAÚCHO	411
	<i>CORPUS</i> DE CONTRASTE DO JORNAL MASSA	411
	<i>CORPUS</i> DE CONTRASTE DO JORNAL ZERO HORA	411
	<i>CORPUS</i> DE CONTRASTE DO JORNAL FOLHA DE SÃO PAULO	411
	<i>CORPUS</i> DE REFERÊNCIA 1 – OBRAS LITERÁRIAS.....	411

1 INTRODUÇÃO

Dizem que não se deve dar o peixe, mas ensinar o povo a pescar. Mas, quando destroçamos seu barco, roubamos sua vara e tiramos seus anzóis, é preciso começar dando-lhes o peixe.
José Mujica

No Brasil, segundo dados do Indicador Nacional de Alfabetismo (INAF), na edição de 2018¹, 88% da população entre 15 e 64 anos apresentam algum problema de compreensão leitora. As razões para essa realidade são muitas e não está no escopo deste trabalho discuti-las. No dizer de Pasqualini (2018, p. 16),

[p]oucos brasileiros entendem o que leem, pelos mais diversos motivos. Assim, enquanto não tenhamos como solucionar e sanar a origem dessa dificuldade em um curto espaço de tempo, é preciso trabalhar, de algum modo, na direção de promover acessibilidade à informação textual para pessoas com escolaridade limitada e pouco hábito de leitura.

É preciso, neste momento, “dar o peixe”, pois a população do Brasil, desde a sua colonização, vem tendo seu barco, sua vara e seus anzóis destroçados por um sistema social extremamente desigual. E este trabalho espera contribuir para que muitos brasileiros recebam o seu peixe, se fortaleçam e construam novos barcos, anzóis, varas e voltem a pescar por si mesmos.

Diante dessa realidade, esta pesquisa toma como pano de fundo as peculiaridades lexicais e sintáticas da linguagem empregada nas Sentenças dos Juizados Especiais Cíveis (doravante JECs) do Poder Judiciário do Estado do Rio Grande do Sul (doravante PJRS). Por terem como objeto a realização do Direito, as Sentenças judiciais são consideradas textos jurídicos, que, como tais, veiculam a linguagem do Direito.

Nesse sentido, cumpre considerar que “o Direito se manifesta através da língua, pois são as palavras que emprega e os enunciados que produz que lhe conferem e confirmam a existência [...]” (Maciel, A., 2001, p. 55). E essa íntima relação entre o Direito e a língua ocorre, segundo Cornu (1990), pelo emprego de certas palavras² com um sentido particular e pela forma como suas proposições são enunciadas. Ao conjunto de palavras com um sentido particular no

¹ **Resultados Preliminares.** INAF, 2018. Disponível em: http://acaoeducativa.org.br/wp-content/uploads/2018/08/Inaf2018_Relat%C3%B3rio-Resultados-Preliminares_v08Ago2018.pdf. Acesso em: 10 dez. 2018.

² ‘Palavra’, neste trabalho, é vista como uma unidade da língua escrita, situada entre dois espaços em branco, ou entre espaço em branco e sinal de pontuação.

Direito, esse autor dá o nome de *vocabulário jurídico*; e os enunciados que o Direito produz, chama de *discurso jurídico*.

Disso se depreende que “a linguagem do direito não está apenas nos termos que emprega, mas também nos textos que produz” (Petri, 2017, p. 43). E esses textos conformariam a linguagem do Direito, que, tal como a da Medicina, Sociologia ou qualquer outra ciência, é considerada uma linguagem especializada³.

Autores como Cornu (1990) e Montoro (1998) afirmam que a linguagem jurídica compreende diversas espécies que se especificam de acordo com a sua finalidade. Abaixo, segue uma lista detalhada dessas espécies segundo Petri (2017, p. 47):

- 1) **linguagem legislativa** – a linguagem dos códigos, das normas; sua finalidade: criar o direito;
- 2) **linguagem judiciária, forense ou processual** - é a linguagem dos processos; sua finalidade: aplicar o direito;
- 3) **linguagem convencional ou contratual** – é a linguagem dos contratos, por meio dos quais se criam direitos e obrigações entre as partes;
- 4) **linguagem doutrinária** – é a linguagem dos mestres, dos doutrinadores, cuja finalidade é explicar os institutos jurídicos, é ensinar o direito;
- 5) **linguagem cartorária ou notarial** – a linguagem jurídica que tem por finalidade registrar os atos de direito.

Considerando que nossa reflexão se centrará nas Sentenças dos JECs do PJRS, estaremos trabalhando com a *linguagem judiciária, forense ou processual*. Entretanto, como tal linguagem é uma subespécie da linguagem jurídica, em nossa pesquisa vamos utilizar a expressão *linguagem jurídica*, por ser mais abrangente.

Nesse cenário, nossa pesquisa pretende refletir sobre como, pela descrição linguística, se poderia colaborar para promover a Acessibilidade Textual e Terminológica (doravante ATT), por meio dos princípios da Linguagem Simples (doravante LS). Assim, trazemos princípios da ATT e da LS para tentar equacionar a relação entre a apresentação das Sentenças dos JECs e as necessidades de compreensão de um público leigo em Direito, o cidadão comum, com escolaridade limitada e pouca experiência em leitura⁴.

Registramos o termo “Acessibilidade Textual e Terminológica” utilizando iniciais maiúsculas, porque, conforme temos observado na mais recente bibliografia, essa temática

³ Seguindo Finatto (2004, p. 342), empregamos o termo *linguagem especializada* e não *linguagem de especialidade*, porque esta última expressão carrega consigo uma ideia de posse da linguagem pelo usuário ou pela área de conhecimento e/ou saber.

⁴ Entendemos “escolaridade limitada” como aquela que vai até, no máximo, o 9º ano do Ensino Fundamental e “pouca experiência em leitura” como pouco contato com material de leitura. Segundo a 5ª edição da pesquisa *Retratos da Leitura no Brasil*, desenvolvida pelo Instituto Pró-Livro, a média anual de leitura no Brasil é de 4,95 livro por habitante. Porém, menos de 3, em média, foram lidos do início ao fim, o que tende a comprometer ainda mais o índice de leitura no Brasil.

surge como um novo tópico de estudos em Terminologia no Brasil. Sob esse tópico e sob a orientação da Prof^a. Dra. Maria José Bocorny Finatto, analisam-se textos, discursos, terminologias, vocabulários e convenções de escrita de diferentes áreas do conhecimento humano tendo em mente subsidiar a facilitação da compreensão por parte de leitores adultos de escolaridade limitada. Nossa pesquisa conduz esse tipo de estudo com textos do âmbito do Direito.

Os Juizados Especiais, ambiente de circulação das Sentenças analisadas nesta pesquisa, foram criados em 1995, com o propósito de simplificar os procedimentos processuais e de fornecer uma resposta mais rápida aos conflitos menos complexos. Conflitos menos complexos não dependem de vistorias, ou perícias, ou outro tipo de prova mais elaborada e têm como valor de causa até 40 salários mínimos vigentes. Com isso, uma das finalidades dos JECs foi a de cumprir “a missão de abrir as portas do Poder Judiciário às pessoas mais carentes, atendendo a uma demanda reprimida, mediante a oferta de um processo rápido, econômico e simples” (NETO, 2006, p.3).

O art. 2º da Lei n. 9.099, de 1995, que criou esses Juizados Especiais, assim estabelece: “O processo orientar-se-á pelos critérios da oralidade, simplicidade, informalidade, economia processual e celeridade, buscando, sempre que possível, a conciliação ou a transação” (BRASIL, 1995).

O princípio da oralidade significa que, na prática dos atos processuais dos JECs, deve prevalecer a comunicação oral: qualquer pessoa pode fazer o seu pedido de forma oral na Secretaria do Juizado. Mas isso não significa que esses atos não sejam documentados por escrito. Os princípios da economia processual e da celeridade buscam a rápida resolução da demanda mediante procedimentos com menor complexidade e menor tempo de tramitação.

E os princípios da simplicidade e da informalidade, segundo Piske (2012), têm por objetivo diminuir a massa dos materiais juntados aos autos do processo.

Em relação à linguagem empregada nas Sentenças dos JECs, parece que nem sempre ela atende a esses princípios do art. 2º da Lei 9.099/95. A título de exemplificação, transcrevemos dois trechos extraídos de Sentenças dos JECs do PJRS. O primeiro, além de apresentar uma sintaxe complexa, chama a atenção pelo vocabulário, que, a nosso ver, seria um fator preponderante para uma eventual falta de compreensão – destacamos algumas palavras no trecho. O segundo, além do vocabulário, constitui uma única frase com 97 palavras e vários níveis de subordinação:

- i) Por fim, **improcede o pleito** relativo aos **lucros cessantes**, pois não há como se evidenciar o prejuízo decorrente do acidente, neste particular. **Deveras, sequer**

há prova segura da profissão da autora, pois a **prova testemunhal** é frágil nesse sentido, **quanto menos** dos valores que deixou de lucrar, **mormente** em se considerando que o único documento **acostado** se refere a um pedido de valor **ínfimo** (fl. 55), de modo a se impor a rejeição deste requerimento. (Processo nº 015/3.14.0000039-0 – RIO GRANDE DO SUL, 2015, grifos nossos.)

- ii) Contudo, creio que tal **cláusula** não serve para afastar o direito da autora na medida em que não foi redigida de forma clara, pois não há prova de que a ré tenha esclarecido à consumidora expressamente que a cirurgia **em exame** estaria **excluída** de atendimento do plano de saúde, sendo que o capítulo que trata das **exclusões** do atendimento são extremamente obscuras, já que não esclarecem quais seriam os procedimentos previstos na referida Tabela INAMP/SUS, não havendo prova de que a consumidora **teve vista** sobre ela, razão pela qual não se vincula a tal **exclusão contratual**. (Processo: 9000378-32.2015.8.21.0027 – RIO GRANDE DO SUL, 2015b, grifos nossos.)

São poucos os profissionais do Direito que relacionam os critérios do art. 2º da Lei n. 9.099/95 com a linguagem, a não ser pelo da oralidade. Entre esses poucos, encontramos Cunha (2018), para quem a simplicidade, além de buscar um menor formalismo, prima por “uma linguagem que preze a clareza em oposição a termos exageradamente técnicos, o que se pode extrair do próprio critério da informalidade”.

1.1 ESTA PESQUISA E OS MOVIMENTOS EM FAVOR DA SIMPLICIDADE NA LINGUAGEM JURÍDICA NO BRASIL

Ampliando o espectro de nossas considerações para o Judiciário brasileiro como um todo, importa primeiro referir que o Poder Judiciário tem por função garantir os direitos individuais, coletivos e sociais e resolver conflitos entre cidadãos, entidades e Estado. E o instrumento para tanto é, em regra, a palavra escrita. É por meio da palavra que o direito se materializa: “O Direito é, por excelência, entre as que mais o sejam, a ciência da palavra. Mais precisamente: do uso dinâmico da palavra” (XAVIER, 2002, p. 1).

Segundo Cornu (1990, p. 15), em cada país, a linguagem jurídica é um uso particular da língua comum. Assim, a linguagem do Direito de um país está intrínseca na língua desse país. E a população desse país, em princípio, deveria ser capaz de compreender o que se encontra registrado nos textos jurídicos em suas diversas formas (petições, recursos, decisões judiciais, etc.). De acordo com Roger Luiz Maciel,

[é] preciso ter em conta que a linguagem jurídica faz parte de um contexto maior, está inserida no âmbito de toda a sociedade. Além dos sagazes profissionais que a ventilam, atinge a massa social que recorre à esfera judicial incessantemente. (MACIEL, R., 2007, p. 3)

Considerando essa realidade, muitos profissionais e instituições relacionados ao Direito têm-se manifestado quanto à necessidade de a linguagem jurídica ser de domínio público, e não

um privilégio de poucos. É o caso da Ministra Ellen Gracie Northfleet, que, em seu discurso de posse na Presidência do Supremo Tribunal Federal, no dia 27-04-2006, assim se manifestou:

Que a sentença seja compreensível a quem apresenta a demanda e se enderece às partes em litígio. A decisão deve ter caráter esclarecedor e didático. Destinatário do nosso trabalho é o cidadão jurisdicionado, não as academias jurídicas, as publicações especializadas ou as instâncias superiores. (NORTHFLEET, 2006, p. 1)

Antes disso, em 2004, a Associação dos Magistrados Brasileiros (AMB) encomendou ao Ibope uma pesquisa sobre a imagem do Judiciário entre os cidadãos brasileiros. Um dos dados extraídos dessa pesquisa foi o de que “a imagem do Judiciário é a de uma caixa preta, misteriosa, pouco acessível ao indivíduo comum e com segredos que apenas os juízes podem decodificar”⁵.

Após essa pesquisa, em 2005, a Associação dos Magistrados Brasileiros (AMB) lançou uma campanha nacional pela simplificação da linguagem do Judiciário. Nessa campanha, a AMB promoveu concursos com vista a premiar estudantes de Direito que desenvolvessem projetos para simplificação da linguagem jurídica e juízes que redigissem suas peças de forma mais simples. Promoveu também palestras com professores de português e editou um manual com explicações sobre o funcionamento da Justiça e com um glossário de expressões jurídicas. Na apresentação desse manual, assim se manifesta a AMB:

A Justiça deve ser compreendida em sua atuação por todos e especialmente por seus destinatários. Compreendida, torna-se ainda mais imprescindível à consolidação do Estado Democrático de Direito. (AMB, 2007, p. 4)

Nessa campanha promovida pela AMB, foi realizado um concurso para premiar os melhores trabalhos relativos à simplificação da linguagem jurídica. Na categoria Magistrados, a Ministra Fátima Nancy Andrighi, classificada em 3º lugar, sugere que o “juiz divulgue, pela internet, textos que expliquem o conteúdo das decisões mais relevantes”. E dá um exemplo que julgamos pertinente transcrever:

Por exemplo, o STJ editou a Súmula 278 que estabelece que “o termo inicial do prazo prescricional, na ação de indenização, é a data em que o segurado teve ciência inequívoca da incapacidade laboral”. Para facilitar a compreensão do referido enunciado por qualquer pessoa, seria interessante que o Tribunal, concomitantemente à edição e publicação da Súmula, divulgasse, pela internet, em seu sítio, o seguinte texto explicativo: “A pessoa, que tiver contratado um seguro de vida e de acidentes pessoais e sofrer um acidente que a torne incapaz para o trabalho, terá um prazo máximo para pedir na justiça o pagamento de indenização. A Súmula 278 do STJ estabelece que esse prazo começa a ser contado a partir do dia em que a pessoa tiver a certeza de que não poderá mais trabalhar” (ANDRIGHI, 2005, p. 1).

⁵ Disponível em: https://www.conjur.com.br/2004-out-13/imagem_judiciario_detalhada_pesquisa_ibope. Acesso em: 10 abr. 2019

O Atlas de Acesso à Justiça⁶ é um exemplo dessas iniciativas. Trata-se de um *site* criado pelo Ministério da Justiça com o objetivo de facilitar o acesso ao Sistema da Justiça, reunindo, para tanto, informações em vídeos e textos. Entre os textos, encontram-se cartilhas jurídicas, que se pressupõe estejam em linguagem mais simplificada, destinadas a esclarecer o cidadão acerca de seus direitos e do funcionamento da Justiça em vários ramos do Direito. Há, ainda, um glossário ABC dos Direitos, em cuja introdução se encontra o trecho “O Direito e o Acesso à Justiça apoiam-se em termos e conceitos cuja compreensão é essencial” (BRASIL, 2016), o que revela preocupação com a linguagem como fator de acesso à Justiça.

Em dezembro de 2019, a Fundação Getúlio Vargas (FGV), em parceria com o Instituto de Pesquisas Sociais, Políticas e Econômicas (IPESPE), concluiu o Estudo da Imagem do Judiciário Brasileiro⁷. Este estudo, tal como o de 2004 realizado pelo Ibope, foi executado também a pedido da AMB. Ele fornece informações sobre como a sociedade brasileira percebe o seu Judiciário e que expectativas alimenta em relação aos seus serviços:

O Estudo, inédito em profundidade, apresenta ainda oportunidades para o Judiciário melhorar a comunicação com os cidadãos e aumentar a aproximação com eles. [...] O Estudo foi realizado no período de agosto de 2018 a dezembro de 2019, com a utilização articulada de metodologias e técnicas qualitativas e quantitativas, abrangendo diversos segmentos de público: a) sociedade (termo usado a seguir para denominar usuários e não usuários dos serviços da Justiça, sendo os jurisdicionados demandantes e demandados); b) advogados e defensores públicos; e c) formadores de opinião. (FGV, 2019, p. 3)

Foram realizadas entrevistas, face a face, com usuários e não usuários da Justiça e entrevistas via chamadas telefônicas com advogados e defensores públicos. Além das entrevistas, em relação às mídias sociais no Brasil, “foram capturados, no período de 01 a 31 de agosto de 2019, 9 milhões de *tweets* sobre o Judiciário e 37 milhões de engajamentos sobre o Judiciário no *Facebook*” (FGV, 2019, p. 9). Foram também levantados dados na mídia internacional:

“foram levantadas postagens de veículos de imprensa em inglês, francês e espanhol, em países da América do Sul (Argentina, Chile, Uruguai), Europa (Reino Unido, França, Espanha, Portugal) e nos Estados Unidos, resultando na análise de 77 órgãos divulgadores no período. (FGV, 2019, p. 9)

⁶ Atlas de Acesso à Justiça. Disponível em:

<http://www.acessojustica.gov.br/pub/sobre/oAtlas/exibirOAtlas.faces> . Acesso em 20 out. 2016.

⁷ Estudo da Imagem do Judiciário Brasileiro. Disponível em: https://www.amb.com.br/wp-content/uploads/2020/04/ESTUDO_DA_IMAGEM_DO_JUDICIARIO_BRASILEIRO_COMPLETO.pdf.

Acesso em: 20 abr. 2020.

Em uma das fases da pesquisa, os respondentes foram solicitados a apontar seu grau de concordância com algumas ideias do senso comum relacionadas à Justiça no Brasil. Entre essas ideias, tem-se a seguinte: *a linguagem jurídica é pouco compreensível*.

O percentual de concordância relativamente a esta afirmação foi de 87%.

No que se refere aos serviços digitais da Justiça, **a linguagem ficou em primeiro lugar entre os fatores negativos**: sociedade em geral (49%), defensores públicos (51%) e advogados (13%).

E, entre as 10 sugestões para melhorar o funcionamento do Judiciário, a relativa à linguagem ficou em 5º lugar, antes mesmo de ações como “simplificar os procedimentos” e “diminuir as custas”.

Mais recentemente, a partir de 2020, surgiram no Brasil vários movimentos em favor da simplificação da linguagem nos órgãos governamentais. Porém, é preciso observar que o Manual de Redação da Presidência da República, que teve sua primeira edição em 1991, já sugeria o uso de uma linguagem simples. É o que vemos a seguir:

- a) Utilizar palavras e expressões simples, em seu sentido comum, salvo quando o texto versar sobre assunto técnico, hipótese em que se utilizará nomenclatura própria da área;
- b) Usar frases curtas, bem estruturadas; apresentar as orações na ordem direta e evitar intercalações excessivas. Em certas ocasiões, para evitar ambiguidade, sugere-se a adoção da ordem inversa da oração. (Manual de Redação da Presidência da República, 2018, p. 20)

Além desse Manual, há muitas outras legislações aqui no Brasil, em todas as esferas de poder, que recomendam a adoção de uma linguagem de fácil compreensão. É o caso, por exemplo, da Lei de Acesso à Informação (Lei n. 12.527/11 – arts. 5º e 8º, inciso I), Lei Geral de Proteção dos Dados (Lei n. 13.709/18 – art. 6º, incisos IV e V), Lei do Governo Digital (Lei n. 14.129/21 – art. 3º, inciso VII). E hoje tramita na Câmara dos Deputados o Projeto de Lei n. 6.256/19, que institui a política nacional de Linguagem Simples nos órgãos e entidades da administração pública direta e indireta. Em 24-05-2022, foi aprovado pela Comissão de Trabalho, de Administração e Serviço Público da Câmara dos Deputados.

No âmbito do Judiciário especificamente, também têm crescido o interesse e os movimentos em favor da adoção de uma linguagem simples. O PJRS, por exemplo, em 2010, lançou o Projeto “Petição 10 Sentença 10”⁸, que “propõe objetividade nas sentenças, o que acarreta sustentabilidade, com economia de tempo e papel”. Segundo Tesheiner (2021, p. 118):

⁸ **Projeto Descomplica**. Disponível em: <https://www.tjrs.jus.br/novo/noticia/conheca-o-projeto-peticao-10-sentenca-10/>. Acesso em: 04 maio 2022.

A proposta, com ampla adesão, é a de se produzir peças mais sucintas, com no máximo 10 páginas, evitando excesso de citações e informações dispensáveis, ganhando, assim, maior presteza na análise das petições, com decisões igualmente objetivas, comprometendo a todos com a celeridade e sustentabilidade.

Em 2020, esse mesmo tribunal lançou, em parceria com o curso de Jornalismo do Centro Universitário Metodista – IPA, o Projeto Simplifica. Esse projeto, em sua primeira etapa, previu a publicação de “10 posts com a tradução de alguns dos mais habituais termos relacionados à Justiça com linguagem coloquial e sob a perspectiva de uma geração vinculada à comunicação digital” (BERTI, 2020). Na continuidade, há previsão da publicação de temas relacionados aos direitos e deveres da cidadania.

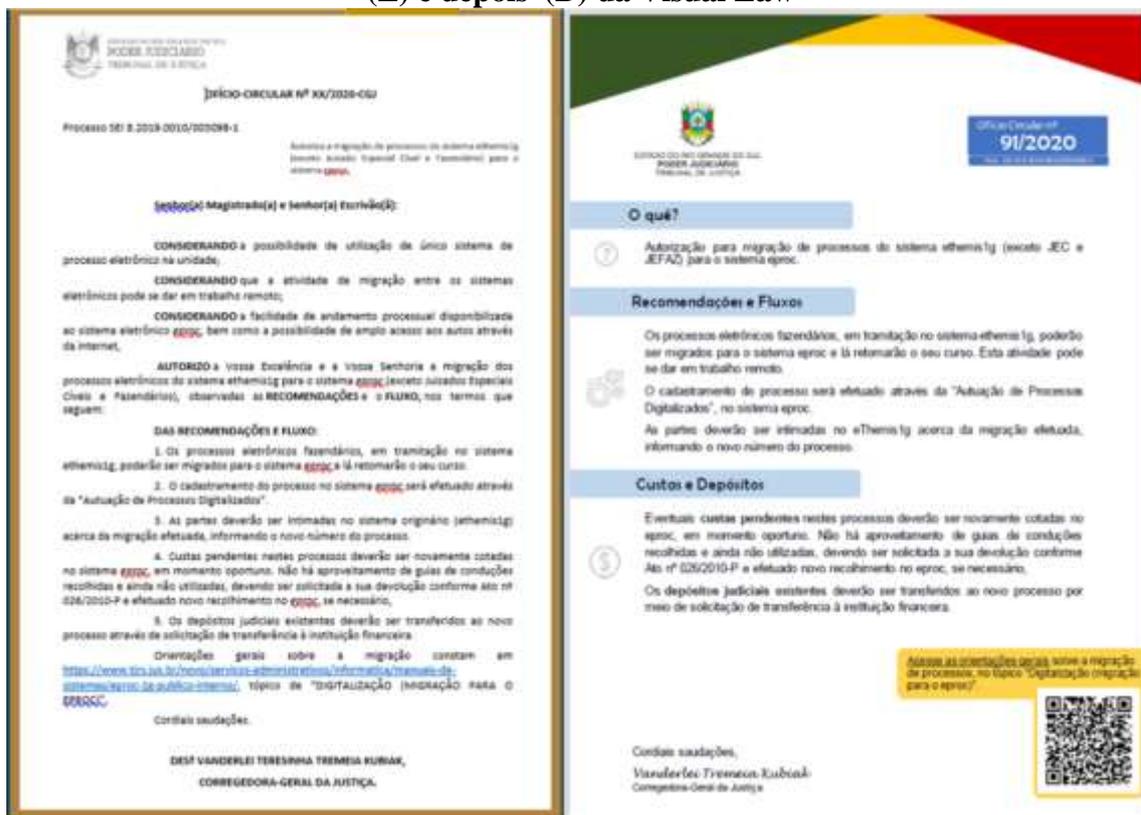
Além desse projeto, a Comissão de Inovação (INOVAJUS) desse tribunal lançou, recentemente, o Projeto Descomplica⁹, que objetiva tornar a comunicação jurídica mais próxima do cidadão:

Além de textos mais curtos, com linguagem usual, a ideia é se valer de recursos visuais, usando técnicas do chamado *Visual Law*. Não estranhe se documentos oficiais começarem a aparecer com cor, gráfico, vídeos, QR Codes fluxogramas e outros elementos visuais que irão ajudar a interpretar o que está escrito. (CAVALHEIRO, 2020).

A Figura 1.1 ilustra esses procedimentos:

⁹ O Projeto Descomplica está incluído no Planejamento Estratégico do TJRS (Portaria n. 106/2020-P).

Figura 1.1 – Exemplo de documento elaborado pela Corregedoria do PJRS antes (E) e depois (D) da Visual Law



Fonte – Site do PJRS. Acesso em: 10 dez. 2021

E bem recentemente, em 25-04-2022, pelo ato Ato nº 024/2022-P, esse tribunal instituiu a política de utilização de Linguagem Simples.

Há ainda, no âmbito do Judiciário brasileiro, outras iniciativas nesse sentido, como a da 6ª Vara Federal da Justiça Federal do Rio Grande do Norte (JFRN), que desenvolveu um modelo de mandado de citação e intimação de penhora com elementos visuais.

Segundo Tesheiner (2021), apesar de caminharem separadamente, Linguagem Simples e Visual Law têm um objetivo em comum: melhorar a transmissão das mensagens. Quanto à inclusão de elementos visuais nos textos, esse autor cita duas iniciativas. Uma delas é do Município de São Paulo, que, no inc. X do art. 4º da Lei nº 17.316, de 2020, estabelece a possibilidade de “usar elementos não textuais, como imagens, tabelas e gráficos de forma complementar”. Outra é a do Laboratório de Inovação Íris, do Governo do Ceará. Segundo as orientações desse órgão, está a de, conforme a necessidade, usar “elementos visuais (ícones, infográficos, fotografias, etc.) para complementar e reforçar a mensagem do texto”¹⁰.

¹⁰ **Conheça a experiência da linguagem simples no Governo do Ceará.** Disponível em: <https://www.ceara.gov.br/2020/08/25/conheca-a-experiencia-da-linguagem-simples-no-governo-do-estado-do-ceara/>. Acesso em: 04 maio/2022.

O Governo de Goiás também lançou, em outubro de 2020, o aplicativo *Tá na Lei*:

A plataforma traz ilustrações para facilitar a identificação do tema e tem uma linguagem acessível. É possível fazer a busca pelo nome do assunto, categoria ou palavra-chave. Além de conferir a legislação na íntegra, o usuário fica sabendo as situações em que deve exigir o que prevê a lei, as penalidades em caso de descumprimento e as formas de denunciar irregularidades. (GOIÁS, 2020) ¹¹

O Tribunal de Justiça do Distrito Federal e dos Territórios, pela Portaria Conjunta nº 91/2021, regulamenta o uso de Linguagem Simples (LS) e de Direito Visual no TJDF:

O TJDF+simples é uma iniciativa institucional que une linguagem simples e direito visual para ampliar o acesso da sociedade à Justiça por meio de comunicações mais claras, acessíveis e inclusivas, conforme disposto na Portaria Conjunta 91/2021 do TJDF. ¹²

Além desses, outros tribunais, como o Tribunal Regional Eleitoral do Paraná e o Tribunal de Contas de Santa Catarina, e também a Advocacia-Geral da União estão adotando, ou estão em vias de adotar, a LS em suas normativas.

Hoje, no Brasil, também existe a Rede Linguagem Simples do Brasil. Foi criada em março de 2021 para ser um espaço público para o debate, o fomento e a construção em torno da Linguagem Simples.

A Rede, em colaboração com a jornalista e especialista em Linguagem Simples Heloisa Fischer, foi fundada por servidoras(es) das organizações a seguir: - Secretaria de Governo Digital do Ministério da Economia do Governo Federal; - Íris | Laboratório de Inovação e Dados do Governo do Ceará; - (011)lab | Laboratório de Inovação em Governo da Prefeitura de São Paulo. ¹³

Há também a Rede Brasileira de Letramento em Saúde (REBRALS). Essa rede reúne em sua grande parte estudiosos e profissionais de saúde que trabalham ou pesquisam sobre como facilitar o compartilhamento e a compreensão de informações relativas à saúde. O grupo existe desde 2013, mas a rede foi oficializada em 2022¹⁴.

Cremos – e esperamos também – que essas iniciativas venham aos poucos a diminuir os tímidos resultados que ainda vemos em relação à acessibilidade da linguagem da Justiça.

¹¹ **'Tá na Lei': aplicativo do Governo de Goiás ajuda cidadão a conhecer as leis e a defender seus direitos.** Disponível em: <https://www.casacivil.go.gov.br/noticias/9153-t%c3%a1-na-lei-aplicativo-do-governo-de-goias>. Acesso em: 05 maio/2022.

¹² **TJDF+simples.** Disponível em: <https://www.auroralab.tjdft.jus.br/tjdft-mais-simples#:~:text=O%20TJDF%2Bsimples%20%C3%A9%20uma,Conjunta%2091%2F2021%20do%20TJDF>. Acesso em: 04 maio 2022

¹³ **Rede Linguagem Simples do Brasil.** Disponível em: <https://redelinguagensimpl.editorx.io/rede-linguagem-br>. Acesso em: 04 maio 2022.

¹⁴ **Rede Brasileira de Letramento em Saúde.** Disponível em: <https://rebrals.com.br/>. Acesso em: 04 maio 2022.

Trazemos aqui uma citação de Luís Fernando Veríssimo, já referida em artigo de minha autoria, MOTTA, 2021:

Nada como acompanhar debates jurídicos para aprender a linguagem esotérica com que os juristas se comunicam. Imaginei que se poderia usar **legendas em português** para o que eles estão dizendo. Mas desconfio que, reduzidos a **uma língua inteligível**, os textos perderiam sua função principal, que é a de nos engambelar. (VERÍSSIMO, 2018, p. 4, *apud* MOTTA, 2021, p. 1-2, grifos nossos)

É uma citação anterior a muitos dos empreendimentos em prol de uma linguagem simples no âmbito governamental e Judiciário citados acima, porém vem a atestar o reconhecimento popular de que os textos jurídicos são, em muitos casos, inacessíveis aos cidadãos brasileiros em geral. Nesse sentido, importa também transcrever o que diz Tesheiner (2021, 117): “E não se engane: mesmo pessoas versadas na área jurídica podem se beneficiar com uma redação clara. Se você alguma vez teve de ler um trecho de uma petição ou de uma sentença mais de uma vez, é porque faltou no texto um pouco de ‘Linguagem Simples’”.

E o que dizer, então, em relação ao cidadão com baixa escolaridade e pouca experiência em leitura?

Souza (2020) afirma, com base no sociólogo Raymundo Faoro, que a herança lusitana do Direito brasileiro tende a fazer com que interesses particulares e patrimoniais controlem o Estado e façam perpetuar um sistema de privilégios no aparelho estatal. Entre esses privilégios, estaria o formalismo jurídico, traduzido por leis retóricas e elegantes. Nas palavras de Faoro (2001, p. 885 *apud* SOUZA, 2020, p. 2): "O poder – a soberania nominalmente popular – tem donos, que não emanam da nação, da sociedade, da plebe ignara e pobre".

Pode-se supor que uma das causas para isso esteja na grande resistência de alguns operadores do Direito, uma vez que esta mudança “acabaria mexendo com a vaidade historicamente construída e intocada de muitos membros desse universo [...]. Existe nostalgia nas longas e eruditas construções gramaticais. O bacharel é ‘treinado’ ao longo de sua preparação acadêmica a dominar o jargão e apreender dos textos as ambiguidades de que se valerá mais tarde como ferramenta de trabalho” (SANTANA, 2017).

É essa também a percepção de Lages (2012, p. 203), para quem iniciativas como a da AMB são louváveis, mas acabam esbarrando na independência dos juízes, que só seguirão a orientação se acharem por bem. Ela cita Souto Maior (2007), juiz e professor da USP, para quem muitos juízes imaginam que a exposição pública fragilizaria a autoridade deles.

Ainda nesse sentido, temos as palavras de Adilson Carvalho:

Por ter consciência da importância do grupo de que fazem parte, aqueles que têm o privilégio de pertencerem ao mundo jurídico fazem de tudo para que esse mundo

sagrado não seja profanado pela presença dos não-iniciados. [...] Nesse processo de violência simbólica que "protege" o mundo jurídico do acesso de grande parte da população, nada é tão eficaz quanto a linguagem jurídica. (CARVALHO, 2017, p. 1)

Diante desse contexto, cremos que nossa pesquisa vem a somar na luta para diminuir o distanciamento promovido pela linguagem jurídica entre a Justiça e o cidadão. Como já mencionado, pretendemos, com base na descrição das peculiaridades lexicais e sintáticas da linguagem empregada nas Sentenças dos JECs do PJRS, verificar o quanto essas peculiaridades tornam, ou não, o texto complexo para o cidadão destinatário das decisões postas nessas Sentenças. Cabe lembrar aqui que muitos desses cidadãos são adultos de escolaridade mais ou menos limitada e com pouca experiência em leitura.

Os resultados das análises realizadas nesta pesquisa poderão subsidiar propostas de simplificação textual. Essas propostas, por sua vez, poderão ser colocadas em um guia de sugestões especialmente dirigido a juristas e a redatores que os auxiliem. Poderão também ser utilizadas para subsidiar a criação de uma ferramenta que “traduza”, em linguagem mais simples, o teor de algumas Sentenças dos JECs do PJRS.

O uso de uma linguagem mais clara e compreensível por um maior número de pessoas possível torna o Direito mais transparente e mais democrático, o que vai ao encontro da consolidação do princípio do estado democrático de direito expresso na Constituição Federal de 1988, em seu art. 1º. Segundo Santana (2017), a

compreensão do direito é necessidade, é transformação urgente e sensata para a sociedade brasileira. A linguagem jurídica, portanto, deve fazer parte dos meios de “luta” pela democratização do conhecimento jurídico: o direito produtor de justiça social deve ser inteligível entre seus atores.

Nesse sentido, cumpre também referir – ainda que seja em outro país, cuja legislação é diferente da do Brasil – o que tem acontecido na Espanha. Segundo Branco (2014), nesse país, o governo se comprometeu a proceder à operacionalização das recomendações da Comissão de Modernização da Linguagem Jurídica – CMLJ (2010) –, para a qual três são os direitos fundamentais dos usuários da Justiça:

o direito de compreender todas as notificações e avisos, pelo que devem conter termos simples; o direito de, nas audiências, serem esclarecidos no que toca à sua incompreensão de termos técnicos; e o direito de, nos julgamentos, a decisão ser declarada com uma sintaxe simples, sem prejuízo do seu rigor técnico, para que esta seja plenamente compreendida pelos seus destinatários. (BRANCO, 2014, p. 29)

Além da Espanha, em Portugal também está havendo movimentos para aproximar a linguagem jurídica do cidadão. Os tribunais de lá estão adotando novos modelos de documentos com uma linguagem mais simplificada. Esses novos modelos fazem parte de um projeto-piloto

iniciado em junho de 2017. Nesse projeto-piloto, verificou-se que os documentos simplificados enviados tiveram um aumento de 68% na adesão de seus destinatários:

[..] quem recebeu o novo modelo percebeu melhor o porquê de estar a receber a notificação, bem como as consequências decorrentes do não pagamento da dívida e da possibilidade de ser intentada uma ação executiva para esse fim.[...] Por outro lado, o novo modelo da notificação deu particular atenção à informação sobre o apoio judiciário, sendo que também neste âmbito os resultados foram significativos – os pedidos de apoios judiciários aumentaram 176% e os casos de apoio judiciário efetivamente concedido aumentaram 142%, o que demonstra que houve mais destinatários das notificações não só a tomar consciência da existência do direito ao apoio judiciário, mas também a beneficiar desse direito. (PORTUGAL, 2021)¹⁵

Considerando situada a nossa pesquisa no âmbito dos movimentos em favor da simplicidade da linguagem jurídica, na próxima seção, procuramos situar nosso trabalho entre os estudos linguísticos no Brasil.

1.2 ESTA PESQUISA E OS ESTUDOS LINGUÍSTICOS NO BRASIL

Além das contribuições para a sociedade de um modo geral, no âmbito das relações jurídicas, este trabalho também pretende contribuir no âmbito dos estudos linguísticos em que se insere. É preciso considerar que, por estudar as relações do Estado com os indivíduos, dos indivíduos com o Estado e dos indivíduos entre si, o Direito é uma ciência de natureza humana, social e normativa. Portanto, a ciência do Direito não se refere a um único domínio de conhecimento.

Quanto à linguagem, em especial, é tema abordado pelos mais variados vieses no meio científico. É o que se pode depreender da seguinte passagem, constante da apresentação do livro *Linguagem e Direito*, produzido pelo grupo de pesquisa Linguagem e Direito, da Universidade Católica de Pernambuco:

Partindo do pressuposto de que o domínio da relação entre a linguagem e a atividade jurisdicional se inscreve numa prática necessariamente transdisciplinar, estuda-se a linguagem como atividade sociocultural e seu funcionamento no Direito. Consideram-se tanto os estudos hermenêuticos de interpretação dogmática quanto as abordagens linguístico-discursivas nas quais se destacam o papel das relações interpessoais, o papel do sujeito produtor do discurso jurídico, as estruturas de participação e os contextos sociais imediatos que interferem nos diversos processos de produção e circulação de sentido das ações linguísticas desenvolvidas no âmbito do funcionamento jurídico. (COLARES, 2010, p. 13)

15 **Tribunais: primeiras notificações simplificadas já estão disponíveis.** Disponível em: <https://justica.gov.pt/noticias/primeiras-notificacoes-simplificadas-ja-estao-disponiveis?> Acesso em: 13 maio 2022.

Esta pesquisa analisa a questão da complexidade textual (doravante CT) nas Sentenças dos JECs do PJRS. Por essa razão, lida com uma das modalidades de ações linguísticas desenvolvidas no âmbito do funcionamento jurídico. E essa modalidade de ação linguística está circunscrita a um determinado domínio do conhecimento humano, o Direito, e, no campo dos estudos linguísticos, pode ser analisada pela ótica da Terminologia¹⁶.

No que se refere aos estudos terminológicos relativos à linguagem do Direito no Brasil, a maioria deles, até o momento, preocupou-se em identificar e delimitar as unidades terminológicas, simples ou complexas, presentes em textos do âmbito jurídico. Os resultados desses estudos possibilitaram a criação de bases de dados, que, por sua vez, serviram de fonte para novas pesquisas e publicações de dicionários terminológicos, tanto unilíngues como multilíngues.

É o caso, por exemplo, do trabalho do Grupo Termisul, da UFRGS, cujo objetivo é a pesquisa teórica e aplicada em Terminologia e Terminografia. Desde sua criação, o Grupo publicou: em 1988, o Dicionário de Direito Ambiental: terminologia das leis do meio ambiente; em 2004, o Glossário Multilíngue de Direito Ambiental Internacional: terminologia dos tratados; em 2006, o Glossário de Gestão Ambiental e, em 2008, a 2ª edição do Dicionário de Direito Ambiental. Depois disso, divulgou o Acervo Termisul e hoje está construindo uma Base de Combinatórias Léxicas Especializadas da Linguagem Legal.

Em outras regiões do país, entre os trabalhos relativos à linguagem jurídica, destacam-se os desenvolvidos pelo grupo de pesquisa Linguagem e Direito, da Faculdade de Direito da Universidade Católica de Pernambuco, sob a Coordenação da Prof^a. Virgínia Colares. O objetivo desse grupo é discutir a construção do discurso jurídico no âmbito da doutrina, da jurisprudência e nas diferentes situações de interação na Justiça.

Entre os trabalhos produzidos pelos integrantes desse grupo, têm-se as publicações de Leonardo Mozdezenski que, desde 2004, refletem sobre as relações entre a linguagem e o Direito. Com base na Análise Crítica do Discurso (ACD) e na Teoria Crítica do Direito, serve-se de cartilhas jurídicas como objeto de estudo e procura investigar as estratégias textuais verbais e não-verbais usadas para divulgar ao público leigo os valores jurídicos considerados certos e objetivos. Esse autor aponta o paradoxo das cartilhas jurídicas, que se explicaria da seguinte forma: de um lado, elas são vistas como uma forma de facilitar o acesso do cidadão à Justiça; por outro, elas podem ser um eficiente mecanismo de manutenção do fosso existente

¹⁶ O termo *terminologia* é polissêmico. Por isso, seguindo Krieger e Finatto (2004, p. 13), grafaremos terminologia com T maiúsculo quando se referir à disciplina e com t minúsculo quando se referir ao conjunto de termos ou unidades terminológicas. Estenderemos tal procedimento à menção de outras disciplinas neste trabalho.

entre o cidadão leigo e o ordenamento jurídico. Esse fosso se dá na medida em que as cartilhas instruem o sujeito a conformar-se com as convenções e relações hegemônicas de poder vigentes, aparentemente atenuadas por certas estratégias discursivas, como o uso de recursos visuais e a tradução da hermética linguagem jurídica para uma linguagem comum.

Destacam-se também os trabalhos do Programa de Pós-Graduação em Inglês e em Estudos da Tradução da Universidade Federal de Santa Catarina, onde em 2013 foi criado o grupo de pesquisa Linguística Forense, cujo objetivo é o estudo da linguagem em contextos forenses, ou seja, em contextos nos quais a linguagem e o sistema jurídico interajam:

Atualmente, o grupo tem realizado pesquisas sobre os seguintes temas: tradução forense, treinamento de tradutor, plágio, perícia forense, linguagem simples, rótulos e avisos e entrevistas policiais. Um dos principais objetivos do grupo consiste em diminuir a distância entre a linguagem e a lei, reduzindo a distância entre os linguistas e profissionais como advogados, juízes, procuradores e policiais.¹⁷ (GRUPO DE LINGUÍSTICA FORENSE, 2016, tradução nossa.)

Entre os temas pesquisados por este grupo de pesquisa, encontram-se os referentes ao *Plain Language*. O objetivo desse movimento, que existe em vários países do mundo, como, por exemplo, Estados Unidos e Inglaterra, é promover uma comunicação entre os governos e seus cidadãos em linguagem clara e acessível:

Uma comunicação é em linguagem simples se a linguagem, a estrutura e a forma são tão claras que o público-alvo pode facilmente encontrar o que precisa, entender o que lê no que encontrou e usar essa informação.¹⁸ (PLAIN LANGUAGE ASSOCIATION INTERNATIONAL (PLAIN), 2017, tradução nossa)

No Brasil, entre os estudos que se debruçam sobre a questão de uma linguagem clara e a consideração do público a que se destina, temos hoje já muitos trabalhos. Entre eles, encontram-se os da jornalista Heloísa Fischer com sua dissertação de mestrado recentemente defendida (FISCHER, 2021) e seu *e-book* (idem, 2018).

Encontram-se também as pesquisas relativas à CT e à ATT, que têm como principal referência as publicações da Prof^a. Dra. Maria José Bocorny Finatto e de seus orientandos. Essa pesquisadora coordenou o Projeto PorPopular¹⁹, cujo objetivo principal foi descrever características do que seria um Português Popular Escrito (PPE). Esse PPE comporia os textos

¹⁷ Currently, the group has conducted research on the following subjects: forensic translation, translator training, plagiarism, forensic expertise, plain language, labels and warnings, and police interviews. One of the group's main objectives is to bridge the gap between language and law, narrowing the distance between linguists and members of the law enforcement professions, such as lawyers, judges, prosecutors, and police officers.

¹⁸ A communication is in plain language if the language, structure, and design are so clear that the intended audience can easily find what they need, understand what they find, and use that information.

¹⁹ **Projeto PorPopular**. Disponível em: < <http://www.ufrgs.br/textecc/porlexbras/porpopular/objetivos.php>>. Acesso em 10 ago. 2018.

de jornais populares cujo público-alvo são leitores das classes C, D e E, com escolaridade correspondente ao Ensino Fundamental completo e pouca experiência em leitura. “Esse novo tipo jornalístico mostra-se como um tipo intermediário, situado entre o jornal tradicional ou de referência, voltado para as camadas mais letradas da população, e o jornal de estilo sensacionalista, um jornal de cunho apelativo” (FINATTO *et al.*, 2013, p. 56)

Em artigo que aborda a questão do letramento científico e a simplificação textual, essa pesquisadora, com base em dados do Ministério da Ciência e Tecnologia e do Indicador de Alfabetismo Funcional, afirma:

Ao que parece, mesmo exposta a textos, livros, seja na internet ou em suporte físico, apenas uma pequena parcela dos brasileiros consegue interagir qualificadamente com a informação que recebe. Poucos entendem o que leem, por vários motivos. [...] Há um desafio de mitigar as desigualdades e promover ações que permitam que o acesso livre às informações também signifique acesso livre ao conhecimento. (FINATTO; EVERS; STEFANI, 2016, p. 138, grifo nosso)

Considerando o trecho por nós grifado na citação acima, cremos que, entre as justificativas para a pesquisa que ora apresentamos, está a de que poderá constituir uma dessas ações que promovam o livre e efetivo acesso ao conhecimento. Assim cremos porque a análise das peculiaridades lexicais e sintáticas da linguagem empregada nas Sentenças dos JECs do PJRS nos dará subsídios para promover a ATT dessas Sentenças para um público leigo em Direito, o cidadão comum, com escolaridade limitada e pouca experiência em leitura.

Embora muitas pessoas reconheçam uma complexidade *a priori* da linguagem jurídica brasileira e a necessidade de ela poder ser compreendida pelo “cidadão jurisdicionado” – no dizer da Ministra Ellen Gracie Northfleet (2006) –, carecemos de estudos linguísticos específicos sobre essa temática no Brasil. Poucos são os trabalhos com o objetivo de descrever e mensurar o uso da linguagem em textos do mundo jurídico, ainda que possam ser citados alguns trabalhos sobre a sua terminologia (MACIEL, A., 2001) ou mesmo guias ou dicionários que visam facilitar a sua compreensão para público leigo.

Em tempos atuais, quando a pesquisa linguística já conta com farto apoio computacional para o tratamento de grandes acervos textuais, especialmente úteis para a descrição de padrões lexicais, vemos favorecido o enfrentamento do tema da complexidade dessa linguagem para o público leigo. Com base nesse enfrentamento descritivo e analítico, conforme acreditamos, será possível apresentar explicações e propor alternativas para diferentes situações-problema, sejam as subjetivamente relatadas pelo cidadão comum, sejam as objetivamente verificadas pelo pesquisador linguista.

Cumpre-nos salientar também que, apesar de nossa análise ser relativa à totalidade do texto, para o seu modo de dizer, os resultados das pesquisas de identificação e delimitação das unidades terminológicas dos textos do mundo jurídico realizadas até aqui, conforme já mencionamos, nos são também úteis. Na medida em que estamos trabalhando com Sentenças judiciais, os resultados desses trabalhos podem auxiliar na identificação e descrição das terminologias presentes nos textos que fazem parte do nosso *corpus* de análise.

Tendo exposto o cenário em que nosso trabalho se insere, passamos a apresentar, no próximo capítulo a delimitação do tema de nosso estudo: hipóteses e questões de pesquisa.

2 TEMA DA PESQUISA

Nosso tema de estudo é a linguagem jurídica no português do Brasil, com foco nas peculiaridades lexicais e sintáticas dos textos de Sentenças judiciais dos JECs do PJRS. Partindo desse tema, essas Sentenças, reunidas em uma amostra de pesquisa, tornam-se o nosso principal objeto de atenção, conformando um *corpus* de estudo.

Assim, com base numa revisão de literatura sobre os temas da complexidade e acessibilidade textual e da facilitação da compreensão de leitura de textos de temática especializada para adultos de escolaridade limitada, buscamos verificar, nesse *corpus* de estudo, a existência de possíveis elementos de complexidade linguística, vista aqui como complexidade textual e lexical-terminológica. A descrição e análise da amostra pretendem reconhecer elementos úteis para que possa ser facilitada, quando necessário, a compreensão de leitura desses textos para pessoas leigas em Direito, com especial atenção para pessoas de escolaridade limitada e pouca experiência em leitura. Nossas unidades de análise centram-se no todo estrutural dos textos das Sentenças, destacando-se o vocabulário empregado, a organização frasal e interfrasal.

Sendo a promoção de uma justiça célere e informal o principal objetivo dos JECs, a pessoa pode buscar o auxílio desses juizados sem a intermediação de um advogado. Se não há a necessidade da intermediação de um advogado, há nessa prática uma pressuposição de que o cidadão será capaz de compreender todos os trâmites do processo e, principalmente, todo o teor da decisão relativa ao seu pedido – a Sentença judicial.

Entretanto, de acordo com o Indicador de Alfabetismo Funcional (INAF), na edição de 2018, apenas 12% dos brasileiros entre 15 e 64 anos atingiram o nível Proficiente. Ou seja, apenas 12 em cada 100 brasileiros conseguem elaborar textos mais complexos, interpretar tabelas e gráficos envolvendo mais de duas variáveis e resolver situações-problema de contextos diversos.

Diante dessa realidade e pelo que já expusemos acerca da complexidade da linguagem jurídica - reconhecida inclusive pelos órgãos relacionados ao Judiciário -, as questões propulsoras desta pesquisa, em termos gerais, foram:

- **Do ponto de vista dos Estudos da Linguagem, do Texto, da Terminologia, da complexidade e acessibilidade textual e da facilitação da compreensão leitora, as Sentenças dos JECs podem ser consideradas textos complexos para leitores adultos de escolaridade limitada e pouca experiência em leitura?**

• Sendo as Sentenças dos JECs consideradas complexas para leitores adultos de escolaridade limitada, o que podem fazer, concretamente, os estudiosos do Texto, da Linguagem, da Terminologia, da complexidade e acessibilidade textual e da facilitação da compreensão leitora em relação à elaboração da escrita dessas Sentenças em prol de uma comunicação dotada de maior ATT?

Inspirados por essas duas grandes questões mais genéricas, estabelecemos o seguinte conjunto de questões específicas de pesquisa.

2.1 QUESTÕES DE PESQUISA

a) Como as Sentenças sob exame se configuram do ponto de vista lexical, morfossintático e terminológico?

b) Como se podem estimar eventuais níveis de dificuldade e/ou de complexidade linguístico-textual dessas Sentenças?

c) Que elementos lexicais, morfossintáticos e terminológicos podem contribuir para que os requisitos legais de simplicidade e informalidade²⁰ expressos no art. 2º da Lei n. 9.099, de 1995, nas Sentenças sob exame, não sejam atendidos?

d) Quais as melhores estratégias de escrita e/ou reescrita para que os textos das Sentenças sob exame possam atender seus requisitos legais de informalidade e simplicidade?

Tendo em conta essas questões, agora postas como itens de investigação, as hipóteses que pretendemos verificar com a pesquisa são as que apresentamos na próxima seção.

2.2 HIPÓTESES DE PESQUISA

Levando em consideração as questões anteriores, estabelecemos um conjunto de hipóteses gerais de trabalho.

HIPÓTESE 1

O texto da maioria das Sentenças dos JECs do PJRS apresenta uma configuração lexical, morfossintática e terminológica pouco coincidente com padrões do Português

²⁰ Tais requisitos foram esclarecidos na introdução deste trabalho.

Popular Escrito (PPE²¹), o que pode dificultar sua inteligibilidade para pessoas de escolaridade limitada e pouca experiência em leitura.

HIPÓTESE 2

Os resultados das pesquisas relativas aos estudos do Texto, da Linguagem e da Terminologia podem subsidiar a elaboração de propostas de escritas das Sentenças dos JECs do PJRS que promovam uma maior ATT a leitores adultos de escolaridade limitada e pouca experiência em leitura.

E, em relação às questões específicas, nossas hipóteses são as que seguem:

Hipótese a

O estudo dos padrões lexicais, morfossintáticos e terminológicos pode auxiliar na caracterização da configuração lexical, morfossintática e terminológica dos textos das Sentenças dos JECs do PJRS.

Hipótese b

A comparação dos textos das Sentenças dos JECs do PJRS com textos que representem o Português Popular Escrito (PPE) e também com textos de literatura dos sécs. XIX e XX, que representam o Português Escrito erudito²², pode ajudar na verificação de eventuais níveis de dificuldade e/ou complexidade linguístico-textual e terminológica dessas Sentenças.

Hipótese c

A presença de palavras e construções sintáticas não muito frequentes nos textos do Português Popular Escrito (PPE), bem como a presença de termos excessivamente técnicos podem contribuir para que os requisitos da simplicidade e informalidade expressos no art. 2º da Lei n. 9.099, de 1995 - relacionados à facilitação e/ou desburocratização no andamento dos processos judiciais – não sejam atendidos.

Hipótese d

²¹ Este PPE comporia os textos de jornais populares cujo público-alvo são leitores das classes C, D e E, com escolaridade correspondente ao Ensino Fundamental completo e pouca experiência de leitura.

²² Português erudito aqui é entendido como aquele que se dá na sua forma mais tradicional, mais culta.

O emprego de palavras e de uma sintaxe mais próximas das que se veem nos textos do Português Popular Escrito (PPE) e a diminuição do uso de termos excessivamente técnicos podem representar estratégias para simplificar o texto das Sentenças sob exame. Com isso, se atenderia aos requisitos da simplicidade e informalidade expressos no art. 2º da Lei n. 9.099, de 1995.

Buscaremos comprovar ou refutar a validade dessas diferentes hipóteses e de seus escopos também por meio da descrição e análise do nosso *corpus* de estudo. Tal descrição e análise é apresentada a partir do Capítulo 6. Nossas questões e hipóteses nos ajudaram a estabelecer nossos objetivos de pesquisa, que relacionamos no próximo capítulo.

3 OBJETIVOS

Diante das nossas hipóteses de investigação, temos os seguintes objetivos gerais:

- a) **Descrever alguns aspectos da constituição lexical, morfossintática e terminológica das Sentenças sob exame frente ao que se verifica em uso no PPE.**
- b) **Propor, com base nos resultados da descrição anterior, estratégias de escrita e/ou reescrita das Sentenças sob exame para ampliar sua ATT para público leigo de escolaridade limitada e pouca experiência em leitura.**

A fim de alcançar esses objetivos, inicialmente descreveremos o nosso *corpus* de estudo para verificar como se configuram os padrões lexicais e sintáticos da linguagem empregada nas Sentenças dos JECs do PJRS: se se aproximam ou se se distanciam dos padrões de uma linguagem cotidiana escrita ou de uma linguagem erudita, como as dos clássicos da literatura dos sécs. XIX e XX. Para isso, seguiremos o seguinte roteiro metodológico:

1) Comparar os padrões lexicais e sintáticos da linguagem empregada nas Sentenças dos JECs do PJRS (*corpus* de estudo – CE) com os de um *corpus* de textos dos jornais Zero Hora e Folha de São Paulo (*Corpus* de Contraste A – CCA), que correspondem a um padrão da linguagem cotidiana escrita dirigida às classes A, B e C.

2) Comparar os padrões lexicais e sintáticos da linguagem empregada nas Sentenças dos JECs do PJRS com os de um *corpus* de textos dos jornais Diário Gaúcho e Massa (*Corpus* de Contraste B – CCB), dirigidos às classes C, D e E. Sua linguagem corresponde a um padrão do PPE.

3) Comparar o léxico empregado nas Sentenças dos JECs do PJRS com o CorPop²³ (PASQUALINI, 2018 – *Corpus* de Contraste CCP), um *corpus* do português popular escrito, que também corresponde a um padrão PPE.

4) Comparar o léxico empregado nas Sentenças dos JECs do PJRS com o *Corpus* de Referência 1 (CR1), composto por obras literárias brasileiras que estão em domínio público²⁴. Por serem obras publicadas entre os anos de 1850 e 1953, apresentam, em tese, uma linguagem mais erudita.

²³ Disponível em <http://www.ufrgs.br/textecc/>.

²⁴ As obras constituintes do CR1 foram extraídas do *site* Domínio Público, disponível em www.dominiopublico.gov.br. Este site, criado em 2004, disponibiliza para o público em geral uma biblioteca virtual, com obras literárias, artísticas e científicas (na forma de textos, sons, imagens e vídeos).

5) Comparar o léxico empregado nas Sentenças dos JECs do PJRS com os do *Corpus ptTenTen: Corpus of the Portuguese Web (Corpus de Referência 2 – CR2)*²⁵, composto por um conjunto de textos coletados da internet e com aproximadamente 4 bilhões de palavras.

6) Propor estratégias de formulação e/ou reescrita de trechos das Sentenças considerados complexos para o nosso perfil de leitor, para aproximar os padrões lexicais e sintáticos da linguagem empregada nas Sentenças dos JECs do PJRS aos padrões do PPE.

7) Verificar, com um grupo de leitores reais e o mais próximo possível do perfil de leitor com pouca escolaridade e experiência em leitura, se as estratégias propostas tendem a favorecer a compreensão das informações pelo público a que se destinam.

8) Reunir os resultados das descrições comparativas e da verificação com os leitores mencionados no item anterior, para propor bases de um guia de sugestões redacionais para juristas e seus assessores, ou para redatores técnicos interessados no tema da comunicação facilitada para pessoas leigas em Direito.

No próximo capítulo, passaremos à exposição dos pressupostos teóricos de que nos valeremos para testar nossas hipóteses e atingir os objetivos propostos.

²⁵ Disponível em <https://old.sketchengine.co.uk/auth/corpora/>

4 PRESSUPOSTOS TEÓRICOS

Como já expusemos na introdução desta pesquisa, os critérios da oralidade, simplicidade, informalidade, economia processual e celeridade da lei que rege os Juizados Especiais nem sempre são seguidos. Na continuidade, observamos que a linguagem poderia ser um dos fatores para tanto e exemplificamos com dois trechos extraídos de Sentenças dos JECs do PJRS. Diante disso, estabelecemos como uma das hipóteses de trabalho o fato de a descrição linguística dos padrões lexicais e sintáticos da linguagem empregada nessas Sentenças poder trazer subsídios para a promoção de uma ATT em relação a elas.

Assim, podemos dizer que estamos lidando com uma das modalidades de ações linguísticas desenvolvidas numa determinada esfera da atividade humana, no caso, o mundo do Direito. Como também já observamos, a relação entre a linguagem e o Direito pode ser analisada pelos mais variados vieses no mundo científico. Em nosso estudo, essa abordagem se dará, no âmbito da Linguística Aplicada (doravante LA), pela ótica dos estudos terminológicos. Contará também com os aportes teóricos da Linguística de *Corpus* (doravante LC), com as ferramentas e métodos da Linguística Computacional (doravante LCOMP) e com os subsídios dos estudos relativos à ATT.

A partir daqui, apresentamos, primeiramente, os pressupostos teóricos relativos aos estudos terminológicos aos quais nos filiamos. Depois, abordamos os princípios da LC. Após, tratamos de questões relativas à Acessibilidade Textual (doravante AT), à CT e à Simplificação Textual (doravante ST). Aspectos relacionados à LCOMP serão tratados na parte referente aos procedimentos metodológicos.

4.1 A TERMINOLOGIA

O nosso objeto de estudo são as Sentenças dos JECs do PJRS, cujos textos veiculam a linguagem empregada no Direito. No campo dos estudos linguísticos, tal linguagem é analisada pela ótica da Terminologia.

Segundo Krieger e Finatto (2004), a palavra *terminologia* é polissêmica, pois tanto pode significar os termos técnico-científicos quanto o campo de estudos. No que se refere ao campo de estudos,

a Terminologia é um campo de conhecimento que vem intensificando os estudos sobre a constituição e o comportamento dos termos, compreendendo desde sua gênese até o exame de suas relações nas mais distintas áreas do conhecimento científico e técnico. Para tanto, encontra subsídios na ciência da linguagem, o que lhe permite avançar no

conhecimento do termo, seu objeto central, bem como daqueles outros elementos que também introduziu em seu quadro de investigação, quais sejam, a fraseologia e a definição. (2004, p.22)

Ainda como campo de estudos, a Terminologia tem sua origem na Teoria Geral da Terminologia (TGT), com os trabalhos de Eugen Wüster nos anos de 1930. Para essa teoria, os elementos essenciais da comunicação profissional são os conceitos e os signos associados a esses conceitos. Os conceitos científicos são “atemporais, paradigmáticos e universais” (KRIEGER; FINATTO, 2004, p. 33), e não variáveis como os significados, que se ativam no contexto discursivo. Por tal concepção, a TGT sofreu críticas de outros estudiosos com posturas mais descritivas, e novas vertentes teóricas surgiram.

Entre essas novas vertentes teóricas posteriores à TGT, nossa pesquisa, por centrar-se não só no léxico, mas também na configuração sintática das Sentenças dos JECs, adotará princípios da Terminologia Textual. Mas outros referenciais teóricos também podem nos ajudar a dar conta do reconhecimento das terminologias.

Um desses referenciais é a Socioterminologia, de François Gaudin (1993), Boulanger (1995) e Gambier (1991). Para essa teoria, o conceito é construído no discurso especializado e é sujeito a modificações devido a variáveis sociais e históricas. Tal postura levou ao reconhecimento da variação terminológica nas comunicações especializadas.

No âmbito do Direito, ainda que os manuais preconizem que o termo a ser empregado é o que consta no texto da lei, no dia a dia forense, verificamos o emprego de muitas variantes, tais como: *precatória* para carta *precatória*; *embargos declaratórios* para *embargos de declaração*; *exordial acusatória* para *denúncia*; *peça inaugural* para *petição inicial*. Salientamos, porém, que, conforme o Manual da AMB (2007), algumas dessas variantes, como o caso de *peça inaugural*, para *petição inicial*, e *exordial acusatória*, para denúncia, parecem configurar o *juridiquês*, linguagem usada mais para elitizar e segregar do que para atender ao caráter técnico e à aplicação da justiça.

A distorção e o uso abusivo do chamado “juridiquês” (emprego de vocábulos de difícil compreensão a fim de elitizar a linguagem jurídica e segregar poder) no discurso jurídico podem trazer consequências irreversíveis à justiça e à sociedade. A “confusão” é notável: juristas teimam, diariamente, em dificultar a compreensão dos textos jurídicos (orais ou escritos) achando que com tal prática, contribuem com a “técnica jurídica” e aplicação da Justiça. Entretanto, com essa prática, só conseguem impressionar Alvazires (juízes), Egrégios Tribunais (nobres tribunais superiores) ou o Excelso Sodalício (Supremo Tribunal Federal). (SANTANA, 2017).

Outro referencial é a Teoria Comunicativa da Terminologia (TCT), de Maria Teresa Cabré (1999). A TCT vê o termo como um elemento de natureza poliédrica: as unidades terminológicas apresentam aspectos linguísticos, cognitivos e sociais concomitantemente.

Como exemplo disso trazemos os casos de *sentença* e *competência*. Tais palavras pertencem ao léxico natural da língua portuguesa, mas, ao serem empregadas em um texto do mundo jurídico, adquirem *status* de termo. *Sentença*, no mundo jurídico, é o texto que representa a decisão do juiz a respeito de um fato que lhe é dado julgar. E *competência*, no mundo jurídico, é termo utilizado para designar os limites de um juízo ou órgão da Justiça. Se um juiz atua numa Vara de Família, não poderá julgar, em regra, questões de acidente de trânsito, de aluguéis, homicídios etc.

Frisamos, porém, que a preocupação central dessas teorias é o termo, ainda que um ambiente textual e discursivo seja levado em consideração. Desse modo, uma posição de análise que privilegia a consideração do texto e dos cenários comunicativos nas áreas especializadas – foco de nossa pesquisa – encaminha-nos para outra vertente teórica: a Terminologia Textual, abordada na próxima seção.

4.1.1 Terminologia Textual

Para Finatto (2004, p. 352), “uma perspectiva de Terminologia textual [...] equivaleria a um estudo do texto que tenha termos e se distanciaria de um estudo de termos em texto”.

Lothar Hoffmann, um dos principais expoentes dessa teoria, entende que “o texto é o signo linguístico primário” (2015, p. 47). Segundo este autor, o texto especializado é instrumento e, ao mesmo tempo, resultado da atividade comunicativa exercida no âmbito de uma determinada especialidade socioproductiva. Além disso, este texto se dá por meio de uma estrutura comunicativa complexa na qual há o autor, com suas intenções comunicativas e as estratégias de comunicação que utiliza, e o destinatário, com suas expectativas.

E as Sentenças dos JECs encaixam-se nessa visão de texto, pois é por meio delas (como instrumento) que se dá uma decisão (como resultado) a um conflito levado aos Juizados Especiais. Além disso, na estrutura comunicativa dessas Sentenças, há o juiz, autor, com intenções e estratégias comunicativas de que se utiliza ao redigir seu texto-sentença, e o destinatário, o cidadão, com a expectativa de ver resolvido o seu conflito.

Por isso, cremos, tal qual Hoffmann afirma, que o texto, e não a palavra ou a frase, é objeto de estudo primordial nas linguagens especializadas. Tal postura, porém, não significa abandonar o estudo do léxico ou da sintaxe. Conforme este autor,

cada uma das unidades linguísticas será considerada, de um modo mais enfatizado, como constituinte de fenômenos relacionados a um nível ou níveis mais altos. [...] É no todo do texto que se pode melhor explicar, funcional e comunicativamente, o uso

linguístico especializado, a preferência por determinados recursos linguísticos. (HOFFMANN, 2015, p. 48)

Essa preferência por determinados recursos linguísticos pode ser ilustrada pela tendência à impessoalização por meio do emprego da terceira pessoa, de procedimentos de indeterminação do sujeito ou pelo emprego de construções passivas, conforme destacamos no exemplo abaixo:

Cuida-se de ação de cobrança fundada em título extrajudicial relativo à venda de mercadorias pela autora ao réu que restou inadimplida. (Recurso Cível, nº 71008531964, Segunda Turma Recursal Cível, Turmas Recursais, [...]. Julgado em: 30-10-2019 – grifos nossos)

Ciapuscio (1998, 2003) igualmente entende que o texto pode ser estudado pelas perspectivas do processo e do produto. Vê o termo como uma unidade léxica empregada preferentemente em áreas de especialidade, mas que pode transcender a esses âmbitos.

No trabalho de 1998, ela analisou textos com a mesma temática dirigidos a leitores com perfis diferentes – especialista, semileigo e leigo. Nos textos destinados a semileigos e leigos, houve muitas operações parafrásticas e variação formal da terminologia. Tais operações demonstraram que fatores de índole funcional e situacional (interlocutores, classe textual, âmbito discursivo) condicionam a seleção, o tratamento e os limites da variação formal e conceitual da terminologia.

Em trabalho posterior, Ciapuscio (2003), propõe uma categorização dos textos científicos em quatro níveis: funcional, situacional, semântico e formal-gramatical.

Nessa direção, Finatto chama a atenção para a contribuição das teorias de enunciação, segundo as quais

no texto científico, assim como em qualquer outro texto em linguagem natural, há, entre outras coisas [...] um sujeito que personifica tanto a voz da ciência [...] quanto a sua própria voz individual, a voz de um cientista ou tecnólogo que tem uma determinada concepção desse saber e da própria interlocução que se estabelece via texto. (FINATTO, 2004, p. 349)

Essa perspectiva de análise pode ser considerada em relação aos tímidos resultados relacionados a uma linguagem mais simplificada e acessível ao cidadão no mundo jurídico. Como uma das causas para isso, estaria a grande resistência dos operadores jurídicos – sujeito enunciador que personifica a voz da ciência do Direito - em adotarem uma nova postura em relação ao seu *modus dicendi*.

Do exposto até aqui sobre as correntes teóricas dos estudos terminológicos, a abordagem que nos parece melhor subsidiar as análises que pretendemos realizar nesta pesquisa é a relativa à Terminologia de perspectiva textual. Assim consideramos, pois nosso olhar, sem desprezar

as unidades terminológicas nele presentes, volta-se, em primeiro lugar, para a totalidade do texto, para a sua configuração lexical e sintática, ou seja, para o seu modo de dizer.

Os resultados advindos das pesquisas de Ciapuscio, por exemplo, poderão servir de subsídios para o estudo a que nos propomos, visto que pretendemos verificar em que medida os textos-sentença por nós analisados se adequam ao seu destinatário, o cidadão comum, em regra leigo nos temas do mundo jurídico. Além disso, no que se refere aos princípios extraídos dos trabalhos de Lothar Hoffmann, o texto é o signo linguístico primário. É em sua totalidade que se pode melhor depreender suas características. É no texto que também se observa, segundo este autor, que “a especificidade das linguagens especializadas se expressa principalmente pela frequência de determinados recursos linguísticos, estatisticamente comprováveis” (FINATTO; KILIAN, 2015, p. 38). Essa noção remete aos trabalhos da LC, cujos postulados teóricos também serão adotados em nossa pesquisa. Na próxima seção, abordamos a Linguística de *Corpus* e suas contribuições para o nosso trabalho, no âmbito teórico e metodológico.

4.2 A LINGUÍSTICA DE *CORPUS*

Segundo Berber Sardinha (2004, p. 3),

[a] Linguística de *Corpus* ocupa-se da coleta e exploração de *corpora*, ou conjuntos de dados linguísticos textuais que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem através de evidências empíricas, extraídas por meio de computador.

Este autor comenta que, antes do advento do computador, também havia *corpora* e cita o *Corpus* Helenístico, criado por Alexandre, na Grécia Antiga. Como exemplo do primeiro *corpus* digitalizado, cita o *Brown Corpus*, que foi criado por Nelson Francis e Henry Kucera em 1964 e que impulsionou muitas pesquisas posteriores nessa área.

Sobre o estatuto da Linguística de *Corpus*, Berber Sardinha (2004) afirma que, devido à possibilidade de os métodos e ferramentas desenvolvidos dentro do âmbito da Linguística de *Corpus* poderem ser aplicados em outras áreas, alguns autores acreditam que ela seja apenas uma metodologia, como, por exemplo, McEnery e Wilson (1996). Entretanto, Berber Sardinha chama a atenção para o que se entende, nesse caso, por metodologia:

Assim, se entendermos metodologia como um modo típico de aplicar um conjunto de pressupostos de caráter teórico, então a Linguística de *Corpus* pode ser entendida como uma metodologia, pois traz consigo algo mais do que simplesmente o instrumental computacional. (BERBER SARDINHA, 2000, p. 356)

Mas ele segue explicando outros pontos de vista acerca da LC. Aponta também que alguns a veem como uma disciplina, com o que Berber Sardinha discorda, pois, para ele, o objeto da LC não é delimitado como em outras áreas. Na continuidade, com base em Hoey (1997) e Leech (1992), Berber Sardinha afirma que a LC seria uma abordagem. Ele faz referência ao termo *abordagem baseada em corpus*, criado por Douglas Biber, influente linguista de *corpus*.

Pasqualini (2018, p. 43) explica a distinção entre abordagem e metodologia na LC da seguinte forma:

Estudos de LC como metodologia, ou *corpus-based*, em geral usam dados de *corpora* para validar, refutar ou refinar uma teoria ou hipótese linguística. Já estudos de LC como abordagem, ou *corpus-driven*, partem do pressuposto de que o *corpus* é fonte de teorizações sobre a língua e compõe, em si mesmo, uma teoria da linguagem. A LC como abordagem sempre será também uma ferramenta metodológica. No entanto, como metodologia, não constitui uma abordagem teórica constituída como subárea independente da Linguística Aplicada.

Conforme a citação anterior, a LC como abordagem é mais abrangente que a LC como metodologia. Por essa perspectiva da abordagem, a pesquisa é conduzida a partir dos dados encontrados no *corpus*. Nas palavras de Zilio (2010, p. 137), “[o] *corpus* não é somente um instrumento de confirmação de dados ou de suposições, mas sim uma fonte destes. E é isso o que mais nos atrai para nos voltarmos a esse tipo de abordagem”.

Dito isso, importa verificar que, ainda segundo Berber Sardinha (2004, p. 38), as pesquisas em Linguística de *Corpus* apresentam algumas características comuns, quais sejam: analisam o uso linguístico em textos naturais - servem-se de coletâneas de textos naturais (*corpus*) para suas análises -, utilizam computadores para essas análises e dependem de técnicas quantitativas e qualitativas. Essas características se baseiam nos dois principais pressupostos da LC: a) fazer uma abordagem empírica; e b) ver a linguagem como um sistema probabilístico.

Uma abordagem empírica, na Linguística, significa dar prioridade aos dados provenientes da observação da linguagem em uso num *corpus*. A língua não é mais vista como um conjunto de estruturas pré-existentes na cabeça de qualquer falante e passíveis de serem acessadas pela introspecção, segundo uma visão racionalista da linguagem. Desenhava-se, com isso, um confronto entre a visão empirista da linguagem – de Halliday (1978) - e a visão racionalista – de Chomsky (1970).

De um lado, Chomsky colocava a linguagem como uma capacidade (competência) do ser humano para produzir um número infinito de declarações com base num número finito de regras (desempenho) *possibilidade*. Para este autor, os processos mentais é que levam ao desempenho linguístico e, por isso, devem ser o foco da pesquisa. Com isso, o estudo do

resultado externo ao processo mental não tem interesse. Ou seja, os *corpora* linguísticos perdem relevância como fonte de pesquisa.

De outro lado, Halliday acreditava numa interdependência entre as características linguísticas e os contextos situacionais de uso da linguagem. Para Halliday, a estrutura social é um aspecto particular da experiência humana e deve ser considerada nos estudos da linguagem. Este autor vê a linguagem como probabilidade, com base no pressuposto de que as características linguísticas não ocorrem de forma aleatória e de que é possível, portanto, detectar e quantificar padrões, regularidades.

Acreditamos que uma abordagem mista, combinando intuição e *corpus*, como já defendeu Kennedy (1998 *apud* SARDINHA, 2000), seja a melhor opção para nos dar suporte nesta investigação. Segundo este autor, o funcionamento da linguagem não pode ser totalmente revelado pelos *corpora* porque eles não permitem a distinção entre estruturas possíveis e estruturas impossíveis. Ou seja, o fato de um determinado elemento não aparecer num *corpus*, mesmo sendo este de uma extensão significativa, não quer dizer que esse elemento não existe. Por outro lado, a ocorrência de uma estrutura em um *corpus* não determina automaticamente sua gramaticalidade. Assim, segundo este autor, o uso de um *corpus* é um dispositivo auxiliar fundamental, especialmente para verificar exemplos e validar intuições.

Especificamente quanto à questão da probabilidade, Berber Sardinha cita as pesquisas de Biber (1988, 1995), que evidenciam que os traços linguísticos variam com relação a textos típicos de contextos comunicativos específicos. É o caso, por exemplo, da alta frequência de adjetivos e substantivos em textos especializados:

O fato de que substantivos e adjetivos constituem entre 50% e 60% do vocabulário dos textos especializados é consequência da necessidade de se ter denominações precisas e específicas para conceitos, objetos e processos especializados, e suas características essenciais. Se contarmos os participios como nomes declináveis com forte tendência à adjetivação, então a barreira dos 60% é ultrapassada. (HOFFMANN, 2015c, p. 218).

Nos textos especializados do mundo jurídico, este fenômeno é facilmente percebido. É o que se vê em expressões como *Direito Penal*, *Direito Civil*, *embargos declaratórios*, *embargos infringentes*, *motivo torpe*, *regime aberto*, *regime fechado*, *recurso adesivo*, *recurso inominado* etc. Em todas elas, o adjetivo está delimitando, especificando o objeto de que se trata e, se for retirado, perde-se a precisão do conceito. Essa caracterização é tão entranhada na linguagem jurídica que, no caso dos Juizados Especiais, mesmo que a Lei n. 9.099/99 não dê nome aos recursos dessa esfera judicial, o operador do Direito, no seu dia a dia, qualificou esses

recursos como *recursos inominados*, numa tentativa de, com essa adjetivação, distinguir os recursos dos Juizados Especiais dos demais que ocorrem no âmbito do Direito.

Nesse sentido, ainda no que se refere à questão da linguagem como sistema probabilístico, outros fatores, como a proficiência linguística do autor do texto, o tipo de texto, a temática, influem na seleção de palavras. E, para verificar essas situações, a observação em dados reais em diferentes contextos de uso é a melhor opção e será a nossa abordagem neste trabalho.

Assim, a comparação que pretendemos fazer nesta pesquisa em diferentes *corpora* – Sentenças judiciais, textos de jornais de referência, textos de jornais populares, o CorPop, o CR1 e o CR2 – levará em conta esses princípios. Ou seja, adotará uma postura empirista, pois analisaremos a linguagem natural sob a forma de *corpora* representantes de registros²⁶ variados, como textos jurídicos, jornalísticos – de referência e populares –, literários e da internet. Além disso, por contarmos com apoio informatizado e trabalharmos com base em dados estatísticos, estaremos lidando com a linguagem como sistema probabilístico.

Tendo em conta que trabalharemos com diferentes *corpora*, importa aqui também verificar quais são os requisitos para que um conjunto de textos possa ser considerado um *corpus*. Conforme a transcrição de Berber Sardinha (2004, p. 3), no início desta seção, *corpora* seriam “conjuntos de dados linguísticos textuais que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade linguística”.

Posteriormente, após algumas considerações de que nem toda coletânea de dados linguísticos pode ser considerada um *corpus* e após análises de algumas definições elaboradas para a palavra *corpus*, Berber Sardinha (2004, p. 18), apresenta como exemplo de definição para *corpus* o seguinte trecho:

Um conjunto de dados linguísticos (pertencentes ao uso oral ou escrito da língua, ou a ambos), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso linguístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise’ (Sanchez, 1995, pp. 8-9)

Em seguida, ele menciona os critérios pelos quais essa definição seria a melhor do seu ponto de vista, quais sejam: a autenticidade (dados linguísticos textuais autênticos, em

26 Utilizamos a expressão “registro”, considerando a seguinte afirmação de Biber (2012, p. 15): “Uso os termos gênero ou registro para fazer referência às categorias de textos definidas de acordo com a situação (como ficção, transmissões desportivas, artigos de psicologia) e tipo de texto para se referir a categorias de texto definida linguisticamente. [...] Embora os registros/gêneros não sejam definidos em termos linguísticos, há diferenças linguísticas estatisticamente importantes entre essas categorias”.

linguagem natural); o propósito (para pesquisa); a composição (textos criteriosamente escolhidos); formatação (possibilidade de serem lidos por computador); representatividade (possibilidade de representar uma língua ou uma variedade dessa língua) e extensão (para ser representativo, tem de ser vasto).

Por representarem critérios que causam algumas controvérsias entre os estudiosos da LC, julgamos pertinente fazer algumas considerações sobre as duas últimas características relativas a um *corpus*: a *representatividade* e a *extensão*.

No que se refere à *representatividade*, Berber Sardinha (2004) explica que, quanto maior o *corpus*, maior será a sua representatividade, porque pode conter mais instâncias de traços linguísticos raros. Ele explica que, como a linguagem é um sistema probabilístico, certos traços são menos frequentes que outros. É o caso de palavras com frequência muito rara, como também de itens lexicais com sentidos menos frequentes. Em um *corpus* maior, há maior chance de essas situações ocorrerem.

Todavia, segundo Biber (2012, p. 12), “o tamanho da amostra não é o fator mais importante na seleção de uma amostra representativa [...]. A representatividade se refere ao quanto uma amostra inclui de toda a gama de variabilidade de uma população”.

Nesse sentido, é o posicionamento de Pasqualini (2018, p. 41-2):

[...] conforme entendemos, as amostras, a especificidade e o tamanho determinam a representatividade do *corpus* conforme a origem dos textos produzidos e coletados. [...] **Assim, um *corpus* pequeno pode ter a mesma validade que um *corpus* grande**, desde que consideradas as suas finalidades. (grifos da autora)

Esta pesquisadora menciona outros autores que também relativizam a questão referente ao tamanho do *corpus* como critério de representatividade, como Lopes *et al.* (2016) e Finatto (2018). Ela cita também Diana Santos (2008), considerada uma referência à frente da Linguateca²⁷, para quem o mais importante é saber o que fazer com o *corpus*, para que finalidade ele será utilizado.

É o que defenderam, também, Aluísio e Almeida (2006). Especificamente quanto à extensão do *corpus*, essas pesquisadoras, citando Sinclair (2005), orientam que o tamanho do *corpus* deve estar de acordo com o tipo de pesquisa a ser realizada:

²⁷ A Linguateca é um centro de recursos distribuído para o processamento da língua portuguesa, www.linguateca.pt, financiado pelo governo português desde maio de 2000, que foi criado após o projecto preparatório intitulado "Processamento computacional do português" (1998-2000). Informalmente considera-se que a Linguateca corresponde ao conjunto dos dois projectos, daí o marcar dez anos da sua existência em 2008, através do livro [Costa, Santos & Cardoso \(2008\)](#). Disponível em: <<http://www.linguateca.pt>>. Acesso em 16-11-2018.

Quando se fala em tamanho de um *corpus*, não se trata somente do número total de palavras (*tokens*) e de palavras diferentes (*types*), mas com quantas categorias (gêneros discursivos, tipos de textos, datas, autores, etc.) um *corpus* deve contar [...]. Para estudos da prosódia, por exemplo, um *corpus* de 100 mil palavras será o suficiente para generalizações com propósitos descritivos; para estudos de muitos processos sintáticos, um *corpus* de 500 mil a 1 milhão de palavras é suficiente; para a criação de dicionários de língua geral, que devem definir os vários significados de suas entradas, gramáticas e usos, seria necessário um *corpus* muito maior, por exemplo, o Bank of English, que apoia a criação de produtos da editora Collins, possui atualmente 530 milhões de palavras. (ALUÍSIO; ALMEIDA, 2006, p. 159)

Ou seja, o que mais importa é a finalidade do *corpus*. E o *corpus* será representativo, na medida em que puder também atender aos objetivos e questões de uma pesquisa: “Além de representativo, o *corpus* deve ser adequado aos interesses do pesquisador, que deve ter uma questão a investigar para a qual necessite de um *corpus* específico” (BERBER SARDINHA, 2004, p. 29).

Beilke (2016), por exemplo, ao compilar um conjunto de *corpora* do pomerano, de uma variedade dialetal cujos falantes encontram-se em processo de diminuição considerável, não pôde distribuir equitativamente os textos. Segundo essa autora,

o “desequilíbrio” (não-balanceamento) do conjunto de *corpora* do pomerano, que nos propomos a compilar, é diretamente proporcional à sua representatividade, ou seja, quanto mais diversificados são os conteúdos e tipologias das amostras, mais representatividade conseguimos obter acerca do pomerano, tendo em vista o objetivo de documentar o léxico e de resgatar a variedade linguística (BEILKE, 2016, p. 83).

Além desses requisitos, Berber Sardinha (2004) afirma que os *corpora* podem ser tipificados com base em critérios como o *modo*, o *tempo*, a *seleção*, o *conteúdo*, a *autoria*, a *disposição interna* e a *finalidade*. Quando descrevermos os nossos *corpora*, na seção relativa aos procedimentos metodológicos, vamos usar como modelo o Quadro 4.1, a seguir, criado com base nessa tipologia.

Quadro 4.1 – Caracterização dos Tipos de *Corpora*

MODO	Falado – Escrito - Ambos	
TEMPO	Sincrônico – Diacrônico – Contemporâneo - Histórico	
SELEÇÃO	De amostragem/Estático	Amostra finita.
	Monitor/Dinâmico	É reciclável para refletir o estado atual de uma língua.
	Equilibrado	Os componentes são distribuídos em quantidades semelhantes.
CONTEÚDO	Língua Geral	Os textos são de língua geral
	Especializado	Os textos são de registros específicos.

	Regional ou dialetal	Os textos representam uma ou mais variedades sociolinguísticas.
	Multilíngue	Os textos são de idiomas diferentes.
AUTORIA	De língua nativa	Os autores são falantes nativos.
	De aprendiz	Os autores não são falantes nativos.
	Autor único	O texto é produzido apenas por um indivíduo.
	Mais de um autor	O texto é produzido por mais de um indivíduo.
	Institucional	O texto é produzido por uma instituição.
FINALIDADE	De estudo	O <i>corpus</i> que se pretende descrever.
	De referência	O <i>corpus</i> , no mínimo cinco vezes maior que o <i>corpus</i> de estudo, é usado para fins de contraste com o <i>corpus</i> de estudo.
	De contraste	O <i>corpus</i> é usado para fins de contraste com o <i>corpus</i> de estudo.
	De treinamento ou teste	O <i>corpus</i> é construído para permitir o desenvolvimento de aplicações e ferramentas de análise.
CODIFICAÇÃO		O texto apresenta informações extratextuais como cabeçalho e etiquetas (morfossintáticas, sintáticas, semânticas, discursivas ou pragmáticas)

Fonte: Produzido pela autora com base em BERBER SARDINHA (2004, p. 20-2); DUARTE (2008, p.161-3).

Importa aqui também mencionar o que muitos pesquisadores destacam em relação ao *corpus*: ele deve ser disponibilizado para futuras pesquisas. Nesse sentido, segundo Monzón (2017, p. 32):

é importante documentar todo o histórico de escolhas para a construção do *corpus*, a fim de que futuros usuários possam ser informados sobre o que esperar de tais dados, em termos de conteúdos e justificativas metodológicas.

Parece-nos também importante tentar estabelecer um cotejo entre os princípios da Linguística de *Corpus* e os da Terminologia Textual, corrente terminológica que embasará nossas análises. Nesse sentido, Zilio (2010, p. 138), assim se manifesta:

Em uma primeira abordagem, vemos que a Terminologia Textual entra com os métodos e aportes teóricos para o trabalho com textos especializados, enquanto a Linguística de *Corpus* entra com o aparato informático, com os pressupostos teóricos e com seu modo peculiar de observar a linguagem.

Na continuidade deste texto, este autor faz outra afirmação que julgamos igualmente pertinente transcrever, porque confere com o que estabelecemos como pressupostos para a presente pesquisa:

É também a partir da observação da linguagem especializada em *corpora* que se percebe mais francamente como a observação de termos é somente um pequeno passo na observação do texto especializado, mesmo que se entenda que esse é um passo importante na caminhada. (ZILIO, 2010, p. 138)

Ou seja, tendo em conta que o texto, segundo Hoffmann (2015a), é o signo linguístico primário, acreditamos que a análise de um conjunto de Sentenças, textos reais, em comparação com outros conjuntos de textos também reais, extraídos de outros contextos de uso, poderá, sim, nos proporcionar um panorama linguístico concreto desses textos.

Entre essas comparações, está a que se encontra no Capítulo 9 deste trabalho, relativa aos aspectos lexicais do CE em relação aos *corpora* de contraste e de referência. Essas análises e as relativas aos aspectos sintáticos, como as que se encontram nos Capítulos 7 e 8, valem-se também dos pressupostos dos estudos estatísticos-lexicais.

No Brasil, a pioneira nos estudos léxico-estatísticos é Maria Tereza Camargo Biderman, para quem “[...] a língua é um código cujos símbolos obedecem a certas frequências determinadas e previsíveis” (1967, p. 119). Assim, se grandes conjuntos de textos são analisados, é possível prever que formas são mais frequentes e, com isso, é também possível melhor verificar as suas peculiaridades.

Essa autora sempre enfatizava que esse tipo de abordagem linguística é tanto quantitativa como qualitativa, não se tratando apenas de mera "contagem de palavras". E tal avaliação é reiterada por Novodvorski e Finatto (2014, p. 15, grifos nossos):

Conforme acreditamos, a LC associou-se a diferentes aventuras de investigação e praticamente nada rejeitou em termos de parcerias de trabalho – o diálogo tem sido uma marca constante, mesmo com aqueles que encaram a LC apenas como um *modus operandi* computacional e quantitativo. A despeito dessa impressão, claro deve ter ficado nesses, pelo menos, primeiros 10 anos de percurso no Brasil, **que vamos muito além de “contar palavras”** e que já prestamos uma contribuição muito importante para toda uma comunidade de pesquisa nacional e globalmente conectada.

Os estudos estatístico-lexicais e os relacionados à Linguística de *Corpus* utilizam alguns termos que empregaremos com bastante frequência neste trabalho. Para que tais termos fiquem claros, a seguir, com base em Silva (2021, p. 62), apresentamos o Quadro 4.2.

Quadro 4.2 – Termos empregados com frequência neste trabalho

Termo	Explicação
<i>Tokens</i>	<p>Palavras, ocorrências ou itens. Cada palavra de uma frase conta como um <i>token</i>, independentemente de ser repetida ou não.</p> <p>Exemplo:</p> <p><i>Minha(1) fatura(2) da(3) luz (4), minha (5) fatura (6) da (7) água (8) e (9) minha (10) fatura (11) da (12) internet(13) vieram (14) com (15) aumento (16) de (17) vinte (18) reais (19).</i></p> <p>(18 <i>tokens</i>, conforme numerais nos parênteses)</p>
<i>Types</i>	<p>Formas ou tipos. Número de palavras diferentes. Quando as palavras são repetidas, elas são contadas apenas uma vez.</p> <p>Exemplo:</p> <p><i>Minha(1) fatura(2) da(3) luz (4), minha (1) fatura (2) da (3) água (5) e (6) minha (1) fatura (2) da (3) internet(7) vieram (8) com (9) aumento (10) de (11) vinte (12) reais (13).</i></p> <p>(13 <i>types</i>, conforme numerais nos parênteses, pois as formas “minha”, fatura” e “da” se repetem)</p>
<i>Palavras lexicais</i> ou <i>Palavras de Conteúdo</i> ou <i>Palavras Plenas</i>	<p>Palavras de conteúdo, ou plenas, ou lexicais, são de classe aberta, como os substantivos, os adjetivos e os verbos (algumas classificações consideram alguns pronomes, advérbios, entre outras classes).</p> <p>Exemplo:</p> <p><i>Minha fatura da luz, minha fatura da água e minha fatura da internet vieram com aumento de vinte reais.</i></p> <p>(Palavras Lexicais: <i>fatura, luz, água, internet, vieram, aumento, vinte, reais</i>)</p>
<i>Palavras gramaticais</i> ou <i>Palavras funcionais</i> ou <i>Palavras Instrumentais</i>	<p>Palavras funcionais ou instrumentais, de classe fechada, como os artigos, as preposições, os pronomes, os advérbios, os numerais e as conjunções (algumas classificações consideram alguns pronomes, verbos auxiliares, entre outras classes).</p> <p>Exemplo:</p> <p><i>Minha fatura da luz, minha fatura da água e minha fatura da internet vieram com aumento de vinte reais.</i></p> <p>(Palavras gramaticais: <i>minha, da, e, com, de</i>)</p>
Lema	<p>Palavras que contêm as suas variações e flexões na sua forma básica ou canônica, chamada lema.</p> <p>Exemplo:</p>

	<p><i>Minha fatura da luz, minha fatura da água e minha fatura da internet vieram com aumento de vinte reais.</i></p> <p>(Lemas: <i>minha, fatura, de, luz, água, internet, vir, com, aumento, vinte, real</i>)</p>
<p>Hapax Legomena</p>	<p>Palavras que ocorrem apenas uma vez no <i>corpus</i>, de única frequência.</p> <p>Exemplo:</p> <p><i>Minha fatura da luz, minha fatura da água e minha fatura da internet vieram com aumento de vinte reais.</i></p> <p>(Hapax legomena – <i>luz, água, e, internet, vieram, com, aumento, de, vinte, reais</i>).</p>

Fonte: Adaptado de BERBER SARDINHA (2004, p. 165-166) *apud* SILVA, 2021.

Outros termos que devem também ficar esclarecidos são **léxico** e **vocabulário**.

Segundo Biderman (1996, p. 32), “*léxico* é o conjunto abstrato das unidades lexicais da língua; *vocabulário* é o conjunto das realizações discursivas dessas mesmas unidades”. Ou seja, o léxico é o conjunto de palavras de uma língua; o vocabulário representa as palavras que são utilizadas pelo falante. Entre as opções do conjunto maior – léxico –, o falante utiliza algumas, que representam o seu vocabulário. E essa escolha reflete “uma dada realidade vivida, uma vez que as palavras são capazes de testemunhar a história e de sintetizar o pensamento humano”. (EVERS, 2018, p. 76).

Até aqui situamos o âmbito de nossa abordagem relativamente aos estudos terminológicos, pela ótica da Terminologia Textual em conjunto com os princípios da LC. Passaremos, agora, a abordar a temática da AT, os conceitos a ela relacionados e a sua relação com a nossa pesquisa.

4.3 A ACESSIBILIDADE TEXTUAL

Quando mencionamos Acessibilidade Textual, pensamos em ideias relacionadas aos termos *acessibilidade* e *texto*.

O primeiro termo – *acessibilidade* – surgiu, segundo Sasaki (1997), na década de 40, para designar a condição de acesso das pessoas com deficiência nos espaços urbanos, nos edifícios e residências e nos meios de transporte coletivo. Porém, atualmente, a noção de *acessibilidade* não é algo apenas associado a espaços físicos. Ela se estende a outros contextos como, por exemplo, o de muitos materiais escritos que, por várias razões, se tornam “ambientes” totalmente inacessíveis a um grande número de pessoas, independentemente de serem portadoras de necessidades especiais ou não.

Quanto ao segundo termo – *texto* –, já discorreremos sobre alguns de seus aspectos quando abordamos a Terminologia Textual, na subseção 4.1.1. Entretanto, retomamos aqui que a palavra *texto* tem origem no latim *textum*, que significa tecido, que, por sua vez, é definido como um conjunto de fios entrelaçados, formando uma trama. Ou seja, um texto pode ser visto como um conjunto de elementos que se conectam de modo a dar forma ao produto (o texto). Essa forma se manifesta por uma estrutura ou sentido. Seja oral, seja escrito, o sentido de um texto só se dará por um processo constante de interação entre autor-texto-leitor/ouvinte.

Segundo Barros (2005, p. 11), um texto é concebido como objeto de significação e como objeto de comunicação. É visto como objeto de significação, porque apresenta uma organização e/ou estruturação que o tornam um ‘todo de sentido’. É considerado objeto de comunicação, porque estabelece uma comunicação entre um destinador e um destinatário. Diante disso, o estudo do texto só é completo se leva em conta tantos os seus mecanismos internos como os fatores contextuais que lhe dão sentido.

Corroborando essa visão, Hoffmann (2015a) afirma que o texto é o *signo linguístico primário*, que a *linguagem se realiza apenas por meio de textos*. Nas palavras deste autor:

O que são lidos, traduzidos, resumidos e trabalhados de diferentes modos são os textos. Todas as outras unidades linguísticas devem ser vistas como seus constituintes, como elementos que mantêm diferentes relações entre si, sem as quais a textualidade não se constitui verdadeiramente. (HOFFMANN, 2015a, p. 47-8)

Sem desconhecer que as considerações teóricas sobre o objeto *texto* são muito mais amplas do que as brevemente apresentadas aqui, passamos a tratar de outros conceitos associados à AT.

4.3.1 Conceitos relacionados à Acessibilidade Textual

A AT está associada a conceitos como *inteligibilidade*, *legibilidade*, *apreensibilidade*, *letramento*, *complexidade textual* e *simplificação textual*. E o que podemos entender sobre esses conceitos?

Segundo DuBay (2004, p. 3), a *inteligibilidade* pode ser vista como a propriedade que torna alguns textos mais simples de ler do que outros. O termo *legibilidade*, por outro lado, refere-se às características físicas do texto, como o tamanho, a forma e a cor de letras, o espaçamento, como os parágrafos estão organizados e elementos da formatação textual, como a diagramação (SILVA, 1985). O termo *apreensibilidade* (que também pode ser uma tradução para a forma inglesa *readability*), por sua vez, está relacionado com a facilidade na

compreensão da leitura, velocidade da leitura e apreensão geral de um texto. Esses fatores relacionam-se com a forma de escrita (prolixa ou sucinta) e com o vocabulário utilizado.

A *complexidade textual*, outro conceito associado à AT, pode ser compreendida como uma propriedade ou condição de um texto relacionada com a demanda de esforço despendido por dado tipo de leitor ou usuário para compreender esse texto.

Essa propriedade seria verificada, pelo leitor, pela presença e combinação de um conjunto de recursos, propriedades ou traços de um dado tipo de escrita. Entre esses recursos, encontra-se o léxico e os tipos de estruturas da superfície do texto, como quantidade de palavras por frase, extensão geral do texto, a variação ou repetição de palavras, a presença de referências pronominais, as anáforas, as elipses, a quantidade de terminologia específica que possa fugir do vocabulário comum para um dado perfil de leitor, entre outros. Acrescentem-se a esses elementos propriedades relacionadas à semântica e à pragmática do texto, tais como o nível necessário de inferências para compreensão de partes ou do todo do texto, pressupostos e subentendidos. (FINATTO; MOTTA, 2019, p. 328-9).

E a *simplificação textual*, por sua vez, seria a concretização da AT, que se materializa pela transformação de um texto considerado complexo – para alguém – num texto mais simples, por meio de uma linguagem e de estruturas sintáticas, em tese, mais adequadas ao leitor-alvo. Esse processo pode se dar de uma forma subjetiva, ou seja, baseado apenas em impressões ou em conhecimentos diversos do redator. Poderá também ser guiado por uma série de procedimentos e de critérios, previamente estabelecidos e mensurados, de acordo com uma metodologia científica. Um exemplo singelo de como se daria na prática seria a utilização de lista de palavras “fáceis” para substituir palavras “difíceis”, como a substituição de “residência” por “casa”, “solver” por “resolver”.

A AT e os conceitos a ela relacionados têm sido objeto de muitos estudos desde o início do séc. XX. Na próxima seção, discorreremos sobre alguns deles.

4.3.2 Estudos referentes à Acessibilidade Textual e aos conceitos a ela relacionados

Pasqualini (2012), em trabalho em que avalia traduções de textos literários nas direções Inglês-Português e Português-Inglês, para confirmar a hipótese de que as traduções para o português eram menos inteligíveis que os textos originais, aborda minuciosamente o tema da CT. Esta pesquisadora informa que a CT é normalmente analisada pela ótica da leitura e do ensino da leitura e apresenta os seguintes encaminhamentos:

- a) Centrar-se em características ou elementos presentes em determinados tipos de textos e associá-los a determinadas dificuldades ou facilidades de compreensão de uma determinada categoria de leitores;

- b) Inferir o *modus operandi* do processamento mental cognitivo de diferentes tipos de leitor mediante aplicações de testes de compreensão após ou durante diferentes tipos de experimentos de leitura com diferentes tipos de textos.
- c) Reunir os indicativos dos itens *a* e *b* em prol da melhoria do ensino de leitura ou da produção de materiais informativos com maior acessibilidade de compreensão por leitores com proficiência de leitura reduzida ou em fase de desenvolvimento. (PASQUALINI, 2012, p. 46)

Esta autora cita vários autores, entre eles Davison e Green (1988), segundo os quais, no âmbito internacional, desde 1920, há registros de pesquisas sobre *readability* ou inteligibilidade ou complexidade linguística. Tais estudos surgiram pela necessidade de adequar materiais de leitura a públicos específicos e analisaram questões relacionadas à compreensão leitora com base em traços textuais mensuráveis. E esses traços textuais mensuráveis normalmente se referem “à dificuldade lexical, baseada na frequência e na extensão das palavras, e a dificuldade imposta pelo tamanho da sentença, a partir do cálculo do número de palavras por sentença” (PASQUALINI, 2012, p. 47).

Entre esses estudos, encontram-se os de Rudolfo Flesch (1911-1986), austríaco que, em 1938, se mudou para os Estados Unidos para fugir do Nazismo. Formado em Direito e Biblioteconomia, ao perceber a dificuldade que muitos imigrantes tinham de compreender materiais escritos em inglês, trabalhou por promover uma linguagem mais acessível a diferentes leitores, de acordo com a sua proficiência. Escreveu e publicou muitos livros com essa temática, entre eles *The Art of Readable Writing* (FLESCH, 1949).

Flesch foi um ativo defensor de uma linguagem acessível (*plain language*) e criador do Índice Flesch de Facilidade de Leitura e cocriador do Índice Flesch-Kincaid de Grau Escolar. O primeiro índice estima o grau de complexidade de um texto com base na relação entre tamanho das frases e tamanho das palavras, e o segundo, desenvolvido em parceria com o cientista J. Peter Kincaid para a Marinha dos Estados Unidos, estima o grau escolar necessário para se compreender determinado texto. Também considera o tamanho das frases e das palavras, mas com pesos diferentes na fórmula.

Tal foi a importância dessas fórmulas elaboradas por Flesch, que foram adaptadas, em 1996, para o português brasileiro, em uma fórmula que compreende as duas. Foi uma adaptação realizada por Teresa B. F. Martins Claudete M. Ghiraldelo, M. Graças V. Nunes e O. N. Oliveira Jr., do Instituto de Ciências Matemáticas e de Computação da USP de São Carlos, do grupo NILC.

A fórmula ficou da seguinte forma:

$$ILF = 164.835 - [1.015 \times (\text{total de palavras}/\text{total de frases})] - [84.6 \times (\text{total de sílabas}/\text{total de palavras})]$$

E o resultado da fórmula pode ser interpretado conforme Tabela 4.1, a seguir.

Tabela 4.1 – Níveis de Leiturabilidade²⁸ segundo a Fórmula do Índice Flesch

RESULTADO	LEITURABILIDADE	GRAU ESCOLAR
100-75	Muito fácil	Ensino fundamental 1 (1º a 5º ano)
75-50	Fácil	Ensino fundamental 2 (6º a 9º ano)
50-25	Difícil	Ensino médio
25-00	Muito difícil	Ensino superior

Fonte: <http://www.ufrgs.br/textecc/ acessibilidade/ autores.php>. Acesso em: 10 nov. 2018.

Os trabalhos de Flesch tornaram-se referência nos Estados Unidos e influenciaram produções editoriais, que passaram a se preocupar com a proficiência de leitura de seus leitores. Portanto, as editoras começaram a produzir materiais graduados por faixas de escolaridade para diferentes tipos de leitores. E até hoje muitos materiais escritos nos Estados Unidos procuram se adequar aos parâmetros estabelecidos por Flesch. Nesse sentido, cumpre referir que vários editores e processadores de texto atualmente contam com o Índice Flesch de Facilidade de Leitura, como, por exemplo, o *Microsoft Office Word* e o Documentos Google.

Porém, conforme já alertaram Davison e Green (1988), além desses traços textuais mensuráveis, como as medições relacionadas ao número de palavras, de sílabas e de frases, outras variáveis devem também ser consideradas. Entre essas variáveis, encontram-se algumas como a complexidade sintática, as características discursivas, a estrutura retórica e o leitor, o seu interesse e o seu objetivo com a leitura.

Nesse sentido, tem-se também os trabalhos de Willian DuBay – mencionado no início desta seção. DuBay foi um padre e ativista norte-americano que, por defender *gays* e os direitos civis nos Estados Unidos, foi expulso da Igreja. Entre esses direitos, ele já colocava o direito de acesso à informação em uma linguagem clara. Após sua expulsão, seguiu divulgando suas ideias em palestras e passou a trabalhar como escritor e professor. Em 2004, publicou o livro *The Principles of Readability*, em que tratou da forma como determinado texto deve se apresentar para atingir seu público-alvo. Algumas das orientações são:

²⁸ *Leiturabilidade* (também chamada de “inteligibilidade”) diz respeito às características do texto que facilitam ou dificultam o ato de ler e compreender para um determinado perfil de leitor.

- Use palavras curtas, simples e familiares ao leitor.
- Evite jargões.
- Use linguagem neutra em relação a qualquer cultura ou gênero.
- Use gramática, pontuação e grafia corretas.
- Use sentenças simples, voz ativa e tempo presente.
- Comece instruções no modo imperativo iniciando as sentenças com um verbo de ação.
- Use elementos gráficos simples, tais como listas de tópicos e passos numerados para tornar a informação visualmente acessível. (DUBAY, 2004, p. 2, tradução nossa)²⁹

Em relação ao Brasil, há os trabalhos pioneiros de Perini (1982) e de Fulgêncio e Liberato (2010), relacionados à facilitação de leitura e ao ensino de língua portuguesa na Educação Básica nacional. Há também os estudos de Kleiman (1997), de Magda Soares (2004), sobre alfabetização e letramento. E, mais recentemente, encontramos as pesquisas de Pereira e Baretta (2018) sobre Leitura e Ensino, assinalando a preocupação de adaptar textos didáticos escolares de acordo com o perfil de diferentes tipos de alunos.

Quanto ao enfoque do assunto da compreensão de leitura para o ensino de línguas estrangeiras, cabe registrar também o pioneirismo de Leffa (1996) ao trazer questões sobre fórmulas e medidas matemáticas de inteligibilidade de textos. Este autor aponta que o processo de compreensão deve levar em conta três aspectos essenciais: o leitor, o texto, e as circunstâncias em que se dá o encontro entre ambos.

Quanto ao leitor, cumpre repisar aqui a observação que fizemos ao delimitar o tema da nossa pesquisa, ou seja, a de que, no Brasil, o Indicador de Alfabetismo Funcional (INAF), na edição de 2018³⁰, aponta que apenas 12% dos brasileiros entre 15 e 64 anos atingiram o nível Proficiente. Ou seja, apenas 12 em cada 100 brasileiros conseguem elaborar textos mais complexos, interpretar tabelas e gráficos envolvendo mais de duas variáveis e resolver situações-problema de contextos diversos.

²⁹ - Use short, simple, familiar words
 - Avoid jargon.
 - Use culture-and-gender-neutral language.
 - Use correct grammar, punctuation, and spelling.
 - Use simple sentences, active voice, and present tense.
 - Begin instructions in the imperative mode by starting sentences with an action verb.
 - Use simple graphic elements such as bulleted lists and numbered steps to make information visually accessible.

³⁰ **Indicador de Alfabetismo Funcional.** Disponível em: http://acaoeducativa.org.br/wp-content/uploads/2018/08/Inaf2018_Relat%C3%B3rio-Resultados-Preliminares_v08Ago2018.pdf. Acesso em: 10 dez. 2018.

Além disso, na relação que estabelecem entre nível de alfabetismo e escolaridade, os dados do INAF apontam que 34% das pessoas com o segundo ciclo do Ensino Fundamental são classificadas como analfabetas funcionais, e 45% alcançam o nível elementar de alfabetismo. Nesse nível, conseguem selecionar uma ou mais unidades de informação, observando certas condições, em textos diversos de extensão média e realizar pequenas inferências. Entre as que completaram o Ensino Médio, apenas 12% demonstraram domínio pleno das habilidades de leitura, escrita e matemática. Já quanto ao Ensino Superior, completo ou incompleto, 34% das pessoas encontram-se no nível pleno de alfabetismo.

Quanto ao segundo aspecto essencial mencionado por Leffa (1996) – o texto –, entram em cena as características textuais já mencionadas: inteligibilidade, legibilidade, apreensibilidade e complexidade textual. Scarton e Aluísio (2010) analisam como essas e outras características textuais podem ser utilizadas para se avaliar a dificuldade ou a facilidade de compreensão de leitura. Elas comentam que, para a compreensão textual, “destacam-se, tradicionalmente, a legibilidade (apresentação gráfica do texto) e a inteligibilidade (uso de palavras frequentes e estruturas sintáticas menos complexas)” (2010, p. 45).

Essas autoras citam alguns exemplos de situações que tornam o texto complexo: “Sentenças longas, com vários níveis de subordinação, cláusulas embutidas (relativas), sentenças na voz passiva, uso da ordem não canônica para os componentes de uma sentença, além do uso de palavras de baixa frequência” (SCARTON; ALUÍSIO, 2010, p. 45). Em relação à macroestrutura do texto, citam alguns fatores facilitadores da compreensão: “organização do texto, coesão, coerência [...] proximidade da anáfora, o uso de marcadores discursivos entre as orações, a preferência por definições explícitas ou a apresentação de informações completas” (SCARTON; ALUÍSIO, 2010, p. 45).

Nesse artigo, as autoras taxam as fórmulas do *Flesch Reading Ease* e do *Flesch-Kincaid Grade Level* de superficiais visto que medem “características superficiais do texto, como o número de palavras em sentenças e o número de letras ou sílabas por palavra” (SCARTON; ALUÍSIO, 2010, p. 45). Dão-lhes, porém, destaque, “pois a primeira é a única métrica de inteligibilidade já adaptada para o português [...] e incorpora o conceito de séries escolares da segunda” (SCARTON; ALUÍSIO, 2010, p. 46).

Quanto ao terceiro aspecto considerado essencial para a compreensão leitora segundo Leffa (1996) – *as circunstâncias em que se dá o encontro entre o leitor e o texto* –, trazemos o tema para o âmbito do texto especializado, enfoque da nossa pesquisa. Na próxima seção, então, abordaremos esta temática.

4.3.3 A Acessibilidade Textual e Terminológica

Finatto (2011), em artigo cujo objetivo foi a reflexão sobre o tema da CT no âmbito dos estudos sobre textos e linguagens especializadas, assinala que, na esfera internacional, as referências são relativamente poucas. Ela explica que talvez isso ocorra porque o trabalho de Davison e Green é de 1920, e o texto científico é institucionalmente visto como tal a partir de 1930, quando ocorreu a primeira reunião internacional de editores de textos científicos. Finatto (2011), porém, cita dois trabalhos dessas autoras dedicados a problemas de compreensão de textos científicos ou técnicos em um nível global. Um deles refere-se ao trabalho em que trechos de manuais de instrução da Marinha da OTAN (BAKER; ATWOOD; DUFFY, 1988) foram apresentados em versões originais e simplificadas para testes de compreensão com um grupo de leitores técnicos de formações diferenciadas. Outro foi um relativo a cartas de *recall* de fabricantes de veículos, cuja proposta era a elaboração dessas cartas de um modo mais acessível para um consumidor leigo.

Sobre essa temática, há também os trabalhos de Ciapuscio (1998, 2003), por nós já mencionados quando discorremos sobre a Terminologia Textual, no item 4.1.1. Os critérios propostos por Ciapuscio (2003) para classificar os textos especializados – funcional, situacional, semântico e formal-gramatical – relacionam-se com a CT, uma vez que contemplam a função ou o propósito do texto e os seus interlocutores. Estes, por sua vez, determinarão os níveis de conteúdo semântico e formal-gramatical.

Entretanto, no que se refere às Sentenças judiciais de um modo geral, de acordo com os dados do INAF sobre o leitor brasileiro que apresentamos, de acordo com o conteúdo semântico e o nível formal-gramatical que esses textos-sentenças apresentam, esse encontro entre leitor e texto, nas circunstâncias de uma decisão judicial que atinge diretamente a vida do cidadão-leitor, nem sempre favorece a compreensão. Quebra-se, assim, a tríade – o leitor, o texto e as circunstâncias em que se dá o encontro entre ambos – considerada essencial por Leffa (1996) para a compreensão leitora. E, com isso, se mantém a distância histórica entre a Justiça e a sociedade.

Como afirmado na introdução deste trabalho, muitos profissionais e instituições ligadas ao Judiciário reconhecem que a linguagem das decisões judiciais deve ser acessível ao cidadão jurisdicionado. Os JECs parecem ter sido criados nessa direção, pois surgiram para *simplificar* a Justiça, e aí se insere a linguagem em que ela é veiculada. Compreendendo que a simplificação da linguagem em que são veiculados os textos da Justiça remetem a ideias de simplificação textual, na próxima seção trataremos desta temática no que se refere ao texto especializado.

4.3.4 A Simplificação Textual e o Texto Especializado

Finatto, Evers e Stefani (2016) propuseram a alunos de tradução do Curso de Letras uma tarefa de simplificar um texto científico em português sobre a Doença de Parkinson (DP). A simplificação, cuja experiência é relatada em um artigo pioneiro que estabelece uma relação entre a simplificação textual, a tradução intralinguística e o letramento científico, tinha como destinatário leitores brasileiros leigos, de baixa escolaridade e com pouco hábito de leitura.

Afirmam as pesquisadoras que a simplificação da linguagem e dos textos escritos ainda é tema polêmico, pois muitos a relacionam com a vulgarização científica. Trazem como exemplo histórico o caso de Lutero quando propôs que a Bíblia fosse transcrita em um alemão que pudesse ser compreendido pelas pessoas na época. Muitos chegaram a questionar se a Bíblia “escrita de acordo com o dialeto baixo alemão do povo germânico de 1600 ainda seria uma fonte verdadeira” (FINATTO; EVERS; STEFANI, 2016, p. 139, grifo das autoras).

Salientam que se trata de uma questão que envolve reflexões filosóficas, epistemológicas e linguísticas, uma vez que se poderia questionar se a reescrita simplificada não deturparia o escrito original e se haveria uma nova forma original para dizer o mesmo. Acrescentam essa reflexão com a citação de Motta-Roth, para quem a popularização de informações científicas em nosso país e o estudo do discurso nesse campo pode influenciar sobremaneira os modos de atuação política de profissionais das áreas humanas e sociais na sociedade. (MOTTA-ROTH, 2009, p. 22).

Ao finalizarem o artigo, assim se manifestam:

De fato, a simplificação textual e a ideia de uma acessibilidade textual e terminológica [...] são caminhos para implantar ações que democratizem o acesso ao conhecimento, dinamizando e transpondo os resultados das pesquisas produzidas no âmbito da academia para a realidade da população brasileira em geral, ajudando-a no seu processo de amadurecimento linguístico. Afinal, simplificar um texto pode, sim, ser uma atitude impulsionadora para que, a partir de um ponto inicial, o indivíduo busque – e consiga – aumentar seus níveis de letramento. (FINATTO; EVERS; STEFANI, 2016, p. 155-6)

Nesse sentido, os movimentos relacionados ao *Plain Language*, já mencionado neste trabalho, trabalham para tornar textos governamentais mais fáceis para a população. Entre esses textos, encontram-se muitos relacionados à medicina e à lei, áreas que afetam a vida da maioria das pessoas. Tais trabalhos inserem-se no que se chama de *health literacy* e *legal literacy*.

Segundo Zariski (2011), a Associação Canadense de Saúde Pública, por exemplo, reconhece que um baixo letramento em saúde pode comprometer a saúde das pessoas da mesma forma que um baixo letramento jurídico pode afetar os direitos legais das pessoas.

Quanto à *legal literacy*, este autor comenta que, num sistema jurídico complexo e com muitas leis e documentos escritos, um letramento básico não é o suficiente para uma efetiva participação no sistema legal. E traz a seguinte referência do Conselho de Tribunais Administrativo do Canadá:

Mesmo que pessoas com baixa escolaridade tenham encontrado uma maneira de lidar com sua rotina diária, elas acham muito difícil ler, entender e usar material relacionado a problemas legais. Elas não entendem os conceitos contidos nas palavras, mesmo que compreendam as palavras isoladamente. Portanto, elas não podem entender o que se espera delas e, muitas vezes, as implicações do que está sendo dito. (Conselho de Tribunais Administrativos do Canadá de 2005, 11 – tradução nossa)
³¹(ZARISKI, 2011, p. 2)

No que se refere à *health literacy*, tem-se a seguinte explicação no site <https://www.plainlanguage.gov/resources/content-types/healthcare/>:

Letramento é mais simplesmente definido como a habilidade de ler e escrever. Nós podemos pensar que o letramento é a capacidade de compreender informações. Nesse contexto, é usual pensar no letramento em saúde como uma habilidade para compreender e comunicar informações sobre saúde. ³² (tradução nossa).

Quanto ao Brasil, há dispositivos legais que trazem orientações relativas à redação semelhantes às apresentadas pelos movimentos relacionados ao *Plain Language* e por DuBay, em seu livro *The Principles of Readability*, já mencionadas neste trabalho. É o caso do Decreto nº 4.176, de 28 de março de 2002, que trata da redação dos atos normativos. Transcrevemos, a seguir, algumas dessas orientações:

Art. 23. As disposições normativas serão redigidas com clareza, precisão e ordem lógica, observado o seguinte:
 I - para a obtenção da clareza:
 a) usar as palavras e as expressões em seu sentido comum, salvo quando a norma versar sobre assunto técnico, hipótese em que se pode empregar a nomenclatura própria da área em que se está legislando;
 b) usar frases curtas e concisas;
 c) construir as orações na ordem direta, evitando preciosismo, neologismo e adjetivações dispensáveis;
 [...]
 II - para a obtenção da precisão:

³¹ Even if people with low literacy have found a way to cope with their daily routine, they find it very difficult to read, understand, and use material related to legal problems. They do not understand the concepts contained in the words, even if they understand the words themselves. Therefore, they cannot understand what is expected of them and often the implications of what is being said. (Council of Canadian Administrative Tribunals 2005, 11)

³² Literacy is defined most simply as the ability to read and write. We can think of literacy as the ability to understand and communicate information. In this context, it is useful to think of health literacy as the ability to understand and communicate health information.

- a) articular a linguagem, técnica ou comum, com clareza, de modo que permita perfeita compreensão do objetivo, do conteúdo e do alcance do ato normativo;
 - b) expressar a ideia, quando repetida no texto, por meio das mesmas palavras, evitando o emprego de sinonímia com propósito meramente estilístico;
 - c) evitar o emprego de expressão ou palavra que confira duplo sentido ao texto; escolher termos que tenham o mesmo sentido e significado na maior parte do território nacional, evitando o uso de expressões locais ou regionais;
- [...].

De qualquer sorte, ainda que haja todas essas orientações, os textos reais do nosso mundo jurídico brasileiro, como também das demais instituições de poder, parecem desconhecer-las e, portanto, carecem de procedimentos que lhes confira uma maior ATT. Nesse sentido, a simplificação da linguagem em que são veiculados os textos da Justiça pode ser um caminho para a concretização da ATT. E, como a ST também pode ser vista como um processo de tradução intralinguística, na próxima seção fazemos algumas considerações sobre essa relação.

4.4 A TRADUÇÃO INTRALINGUÍSTICA E A SIMPLIFICAÇÃO TEXTUAL

No capítulo introdutório desta pesquisa, transcrevemos um trecho de uma crônica de Luís Fernando Veríssimo em que este autor, por sua manifestação, atesta o reconhecimento popular de que os textos jurídicos precisam, em muitos casos, ser traduzidos para tornarem-se acessíveis aos cidadãos brasileiros em geral.

Especificamente em relação à tradução, Jakobson (1959) a considera um fenômeno que ocorre em todas as transações linguísticas e que pode ser dividido em três tipos: tradução intralinguística ou reformulação; tradução interlinguística ou tradução propriamente dita; e tradução intersemiótica. Embora os estudiosos da tradução mencionem e reconheçam esses tipos de tradução de Jakobson, a tradução intralinguística e a tradução intersemiótica são frequentemente classificadas como periféricas ou de não muita relevância para a disciplina de estudos de tradução.

E é com base na percepção da pouca relevância dada à tradução intralinguística entre os estudos da tradução que Zethsen (2009) pretende descrever a tradução intralinguística e as estratégias nela envolvidas. Essa autora chama a atenção para as relações sociais modernas que estão a exigir atividades de tradução que vão além dos limites do que uma Teoria da Tradução tradicional consideraria como tradução. É o caso de traduções funcionais, como as relativas à comunicação entre especialista e leigo, traduções de livros para crianças ou para versões mais adequadas a um determinado público-alvo, legendas para surdos, etc.

Para este estudo, Zethsen comparou 4 versões dinamarquesas de cinco versos do Capítulo 2 do Evangelho de Lucas, da Bíblia. Ressaltando o fato de a Bíblia ser um livro que já sofreu muitas traduções interlinguísticas, diz que o seu interesse foi verificar como a mesma mensagem seria reformulada para atender a diferentes públicos-alvo. Assim, com base nesse estudo de caso e na revisão de literatura a que procedeu, ela menciona que os principais fatores envolvidos na tradução intralinguística são: conhecimento, tempo, cultura e espaço. E acrescenta, ainda, que a *simplificação*, impulsionada por esses quatro fatores, seria a palavra-chave na tradução intralinguística.

Assim, segundo Zethsen (2009):

- *Em relação ao conhecimento*, que foca na habilidade de compreensão do público-alvo, no seu conhecimento prévio ou nível de conhecimento do tema do texto, a simplificação se daria com o emprego de palavras e sintaxe mais simples.

- *Em relação ao tempo*, que se refere a casos em que a distância temporal requer uma nova versão, a simplificação ocorreria com o emprego de palavras e estruturas sintáticas mais contemporâneas ao momento em que ocorre a tradução intralinguística.

- *Em relação à cultura*, que se refere à necessidade de explicar referências culturais em um texto em que o tempo ou conhecimento geral de antecedentes impedem que o grupo-alvo entenda, mesmo que as línguas envolvidas sejam as mesmas, a simplificação ocorreria com o emprego de expressões mais próximas da cultura do público-alvo.

- *Em relação ao espaço*, que remete a situações de redução ou de extensão do texto, a simplificação se daria com a retirada de informações, com o acréscimo de uma explicação para tornar a informação mais acessível ou, então, com a reestruturação de todo o texto.

Finatto e Tcacenco (2021), em trabalho mais recente, revisam o tema da tradução do ponto de vista de vários autores e situam a simplificação de textos especializados para leigos como um tipo de tradução intralinguística. Reforçam que a simplificação textual e/ou tradução intralinguística tem por fim a construção de um texto acessível. Nas palavras desses autores:

Um texto será acessível, no seu todo, se apresentar informações de forma clara não só para um público específico, conseguindo prever diferentes diálogos. Essa condição aplica-se mesmo a artigos científicos, que muitas vezes precisam ultrapassar um direcionamento muito pontual. É o caso, por exemplo, nestes tempos de pandemia, de um artigo sobre COVID-19, escrito por um infectologista, que precisará ser (bem) entendido por um historiador e que receberá uma ‘tradução simplificada’, feita por um jornalista. Prever ‘outros’ leitores e acolhê-los, já desde a primeira escrita do texto, é promover a ATT. (FINATTO; TCACENCO, 2021, p. 33-34)

Depreende-se desses fatores que, na tradução intralinguística, o que mais importa são as necessidades do público-alvo, e não apenas o que se coloca no texto de origem. No Estudo-

Piloto 2, relatado na continuidade deste trabalho, e no Capítulo 10, servimo-nos desse recurso de tradução intralinguística para tentar simplificar alguns trechos de sentenças judiciais e também uma Sentença no seu todo, de modo a atingir o nosso leitor-alvo: um cidadão com pouca escolaridade e pouca experiência em leitura.

Antes disso, na próxima seção, vamos apresentar alguns estudos relacionados à temática da ATT que têm sido desenvolvidos, como já referimos brevemente no Capítulo 1.

4.5 ESTUDOS RELACIONADOS A ESTA PESQUISA

O tema da ATT, desde 2016, tem sido objeto de estudo por parte de um grupo de investigações liderado pela Prof^a. Finatto. Isso tem sido feito junto à linha de pesquisa “Lexicografia, Terminologia e Tradução: Relações Textuais” do Programa de Pós-Graduação da Universidade Federal do Rio Grande do Sul (PPG-LETRAS-UFRGS).

Sob o tópico da ATT, analisamos textos, discursos, convenções de escrita, terminologias e vocabulários de diferentes áreas do conhecimento humano tendo em mente subsidiar a facilitação da compreensão por parte de leitores adultos de escolaridade limitada. Como protótipo de leitor, fixamos a figura de um cidadão brasileiro adulto, com idade entre 25 e 50 anos, trabalhador ou trabalhadora das classes socioeconômicas C e D, de pouco letramento e com pouca experiência em leitura, cuja escolaridade é equivalente ao Ensino Fundamental completo. (FINATTO; MOTTA, 2019, p. 319)

Os macroprojetos de pesquisa em desenvolvimento nesse grupo são:

- a) *Acessibilidade Textual - Da Doença de Parkinson a cuidados básicos em Pediatria: acessibilidade textual e terminológica para leitores brasileiros de baixa escolaridade;*
- b) *Fundamentos Linguísticos para a Acessibilidade a Informação Científica Para Leitores Adultos de Escolaridade Limitada: simplificação textual, gramatical, lexical e terminológica em ciências da saúde.*

Em torno desses projetos “guarda-chuvas”, desde 2016, associaram-se os seguintes estudos:

- a) *A Linguagem do Patrimônio Cultural Brasileiro: Conservação dos Bens Culturais Móveis – Acessibilidade textual e terminológica - Iniciação Científica; junto ao Grupo TERMISUL21F³³.*
- b) *Recuperação da Informação e Representação do Conhecimento em Bases de Textos Científicos de Linguística e de Medicina: Padrões e Processamento Automático da Linguagem - Iniciação Científica.*

³³ Projeto Terminológico Cone Sul. Disponível em: www.ufrgs.br/termisul.

- c) *Da Doença de Parkinson a Cuidados Básicos em Pediatria: Acessibilidade Textual e Terminológica para Leitores Brasileiros de Baixa Escolaridade* - Iniciação Científica.
- d) *Abaixando o cocho: adaptação de textos sobre doenças causadas pela inalação de amianto destinados para o público leigo* - trabalho de conclusão de curso de graduação (CARPIO, 2017).
- e) *Complexidade Textual e Terminológica em Língua Portuguesa: da Agronomia para os Agricultores* - mestrado concluído (FETTER, 2017).
- f) *Textos de Divulgação para Leigos sobre o Transtorno do Estresse Pós-Traumático em Português: alternativas para a acessibilidade textual e terminológica* - mestrado concluído (SILVA, 2018).
- g) *Tradução Especializada Acessível (TEA): revisão do tema e proposta de disciplina para cursos de graduação em tradução* – mestrado concluído (PARAGASSU, 2018).
- h) *Complexidade Textual em Língua Portuguesa: simplificação e textos institucionais para o cidadão* - tese de doutorado concluída (PASQUALINI, 2018).

Entre esses estudos, cumpre mencionar o de Carpio (2017). Esse estudo gerou o trabalho de conclusão³⁴ de curso de graduação (TCC) com textos sobre doenças respiratórias associadas ao trabalho. Nele, a autora descreveu e analisou textos elaborados pelo Ministério da Saúde (MS) e dirigidos ao público leigo. Tais textos versam sobre o amianto e as doenças provocadas pela inalação desse produto. O trabalho incluiu a medição do grau de legibilidade dos textos, a determinação do perfil de leitura do público-alvo e a avaliação dos níveis lexical, estrutural e sintático do *corpus* de estudo. Os resultados mostraram que os textos analisados apresentam características que os tornam inadequados ao seu leitor-alvo, os trabalhadores que lidam com o amianto.

Além dessa etapa descritiva, Carpio (2017), com base nas características potencializadoras de CT observadas, apresentou propostas de reescrita mais simples e, em tese, mais afeitas à compreensão desses trabalhadores. As alterações apresentadas incluíram, além de aspectos linguísticos e terminológicos, recomendações quanto ao ordenamento da informação indicados por uma médica especialista que trata de pacientes acometidos por Pneumopatias Ocupacionais na Santa Casa de Misericórdia de Porto Alegre - RS.

Em sua dissertação de Mestrado, Fetter (2017) descreveu a apresentação de terminologias, de acordo com padrões oracionais, em textos divulgativos e educacionais de instituições de assistência agropecuária destinados a agricultores familiares brasileiros. Tais

³⁴ Trabalho desenvolvido no âmbito do projeto “Recuperação da informação em representação do conhecimento em bases de textos científicos de Linguística e de Medicina: padrões e processamento automático da linguagem”, financiado pelo Programa DOC-FIX, FAPERGS-CAPES, entre 2012 e 2016, Edital 09/2012, coordenado pela Profa. Dra. Maria José Bocorny Finatto, com a participação da Profa. Dra. Alena Ciulla, que orientou o TCC citado.

agricultores normalmente apresentam escolaridade formal bastante limitada. Com base nos pressupostos teóricos da Terminologia de perspectiva textual (FINATTO, 2004) e da Linguística Sistêmico-Funcional, Fetter (2017) analisou 4.850 orações de 30 folhetos da Associação Riograndense de Empreendimentos de Assistência Técnica e Extensão Rural (EMATER-RS) e de 30 folhetos da Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA).

Após esse exame, a pesquisa apontou alguns elementos que potencializam a CT desses materiais, especialmente o modo de apresentação frasal e as naturezas das terminologias empregadas. Também apresentou, ainda que muito resumidamente, algumas alternativas de reescrita teoricamente mais úteis para uma ampliação da ATT desses materiais.

Outro trabalho a ser mencionado é a pesquisa de Silva (2018). Nessa pesquisa, após descrever e analisar a linguagem especializada associada ao tema do Transtorno do Estresse Pós-Traumático (TEPT) em 10 artigos de divulgação científica em português para leigos, o autor propõe oito diferentes propostas de reescritas para cada um dos textos. O objetivo de cada proposta foi colocar “em teste”, via medições de elementos potencializadores de CT, um determinado tipo de estratégia de apresentação do texto. As medições testadas foram:

- a) Índice Flesch (IF);
- b) análise semântica latente (ASL)³⁵;
- c) relação *type-token* (TT)³⁶;
- d) densidade semântica (DS)³⁷;
- e) incidência de substantivos (S);
- f) incidência de verbos (V);
- g) incidência de adjetivos (ADJ);
- h) incidência de advérbios (ADV); e
- i) incidência de pronomes (P).

Ao comparar o desempenho das medidas, Silva (2018) concluiu que as três mais relevantes e com melhores resultados em termos de redução da CT foram (em ordem): simplificação lexical, redução de adjetivos e redução de informação.

³⁵ A ASL explora as relações semânticas implícitas no texto por meio de uma análise de distribuição e de associação recorrente entre palavras. Além de considerar palavras isoladas, leva em conta as combinações e usos próximos, destacando relações entre as palavras de um texto, tais como “A é parte de B” ou “A é um tipo de B”. Essa é uma técnica matemática de indexação de conteúdos de textos muito usada em buscadores de texto disponíveis na internet, e é uma técnica das pesquisas de PLN, as quais naturalmente, têm uma visão particular de Semântica (GEAN; KAESTNER, 2003).

³⁶ Relação entre o número de palavras diferentes (*types*) e o número de total de palavras (*tokens*). Quanto maior for o número de *types*, maior será a riqueza e a variedade do vocabulário. Quanto menor for o número, mais repetitivo é o texto.

³⁷ A densidade semântica considera o número de palavras de conteúdo (lexicais – substantivo, adjetivo, advérbios e verbos) dividido pelo número de palavras funcionais (gramaticais – preposições, conjunções, numerais e artigos).

Paraguassu (2018), em sua pesquisa de mestrado, propõe a criação de uma disciplina ou módulo de estudos especialmente dedicado aos temas da ATT em meio à grade curricular do curso Bacharelado em Letras-Tradução da UFRGS. A pesquisa é fruto de um trabalho realizado com alunos de Tradução, que foram solicitados a simplificar textos institucionais em português sobre a Doença de Parkinson para um leitor de escolaridade limitada.

A relevância deste trabalho está na discussão relativa ao trinômio Complexidade Textual, Simplificação Textual e Acessibilidade Textual no âmbito da formação universitária de tradutores e no acréscimo de conhecimento e valor à formação de tradutores que já atuam ou que atuarão no mercado de trabalho como tradutores, revisores e redatores.

Além desses estudos, há também a pesquisa de doutorado de Pasqualini (2018). Ainda que esta pesquisa não trabalhe com textos especializados, o resultado desse estudo pode ser muito útil para a tarefa de simplificação de textos de qualquer natureza para leitores brasileiros adultos. Essa pesquisa gerou o CorPop (PASQUALINI, 2018), que, conforme já mencionamos no Capítulo 3, é um *corpus* de referência do português popular escrito.

Tendo em vista que este *corpus* faz parte do conjunto de *corpora* a serem utilizados nesta pesquisa, maiores informações a seu respeito são colocadas na seção 5.1.4.

Outro trabalho no âmbito das investigações coordenadas pela Prof^a. Finatto é uma pesquisa voltada ao aspecto histórico da ATT. Trata-se de um estudo realizado por esta pesquisadora em textos escritos pelo médico João Curvo Semedo, no séc. XVIII, um dos mais famosos do seu tempo em Portugal. Numa época em que o latim era a língua dos textos científicos, ele optou por escrever seus manuais de Medicina em português. Além disso, buscava facilitar a compreensão por parte dos profissionais da Saúde daquela época, especialmente os que não conhecessem o latim. A ideia dessa iniciativa de estudo histórico sobre a ATT é organizar um *corpus* histórico como uma amostra composta por obras “facilitadas”, relacionadas ao tema “doenças e seus tratamentos”. Esse *corpus* está sendo planejado para aproveitamento em atividades didáticas associadas à Terminologia Diacrônica ou sócio-histórica, à Lexicografia e à Linguística Histórica.

Ainda no tema relativo à Saúde, esse grupo tem desenvolvido estudos-pilotos pontuais, com a participação de estudantes de Graduação e de Doutorado, em textos do Ministério da Saúde do Brasil (MS) sobre o tema da sífilis, sarampo e vacinação. Esses materiais, em regra, têm sido examinados em comparação com materiais institucionais sobre a Doença de Parkinson e com suas respectivas facilitações para leigos de pouca escolaridade, tema dos macroprojetos desse grupo de pesquisa.

Nesse contexto insere-se a pesquisa que ora apresentamos. Ainda que a nossa temática não se relacione diretamente com os temas de Saúde Pública, os resultados dele advindos poderão contribuir na ponderação das necessidades do nosso protótipo de leitor-alvo antes citado. O cenário, a temática e o gênero textual podem variar - do Direito à Medicina -, mas apresentam elementos em comum: a figura de um mesmo leitor-destinatário, o brasileiro adulto de escolaridade limitada, e as suas necessidades de entendimento de informações importantes para a sua vida como cidadão.

Na próxima seção, então, passaremos a detalhar os procedimentos metodológicos adotados em nossa pesquisa.

5 MATERIAIS E MÉTODOS

Para que possamos situar as especificidades textuais e léxico-gramaticais dos textos com que lidamos, seguindo as orientações da Linguística de *Corpus* e os preceitos de uma Terminologia de perspectiva textual, torna-se necessário realizar uma série de contrastes e comparações com outros textos. Este capítulo visa situar a compilação do nosso *corpus* de estudo e situar os materiais e instrumentos para as devidas comparações e contrastes.

Assim, realizaremos, nesta pesquisa, um estudo quantitativo e qualitativo sobre os padrões lexicais e sintáticos de um *corpus* de Sentenças dos JECs do PJRS (*corpus* de estudo - CE) – caracterizado na seção 5.1.1 -, em contraste com os seguintes *corpora*:

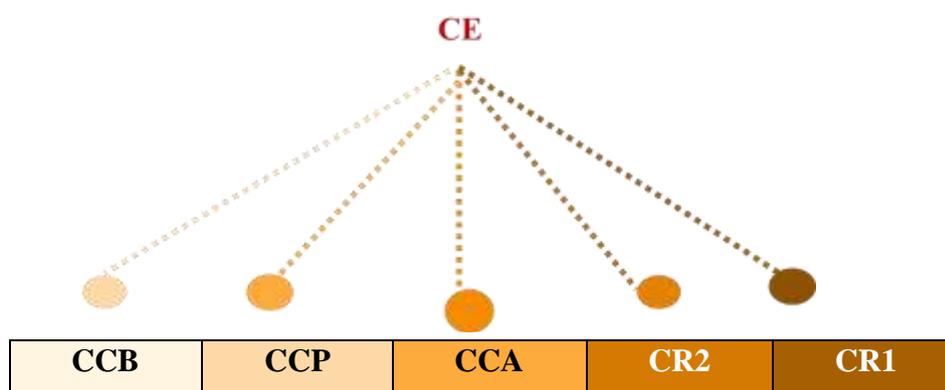
- a) *Corpus* de textos dos jornais Zero Hora e Folha de São Paulo (*Corpus* de Contraste A – CCA), que correspondem a um padrão da linguagem cotidiana escrita culta.
- b) *Corpus* de textos dos jornais Diário Gaúcho e Massa (*Corpus* de Contraste B – CCB), que correspondem a um padrão da linguagem popular escrita simples, o Português Popular Escrito (PPE).
- c) CorPop (*corpus* de contraste – CCP), composto por textos jornalísticos e literários com base no nível de letramento médio dos leitores do país e representativos do PPE.
- d) *Corpus* de Referência 1 (CR1), composto por obras da literatura brasileira (de 1840 a 1956) que já estão em domínio público e correspondem a um padrão de linguagem escrita culta não atual.
- e) O *Corpus* ptTenTen: *Corpus of the Portuguese Web* (*Corpus* de Referência 2 - CR2), composto por um conjunto de textos coletados da internet, que correspondem ao padrão de linguagem cotidiana escrita culta.

Como já mencionamos, esta pesquisa serve-se dos pressupostos teóricos da Linguística de *Corpus*, que preconiza que somente por meio de comparações e contrastes é possível delimitar o que é verdadeiramente peculiar em um *corpus* de linguagem específico. O objetivo é “fornecer uma norma com a qual se fará a comparação das frequências do *corpus* de estudo” (SARDINHA, 2004, p. 97).

O que propomos, então, é um estudo comparativo entre esses cinco *corpora*, a fim de verificar o quanto os padrões lexicais do *corpus* de estudo (CE) são próximos ou distantes dos padrões lexicais dos textos dos *corpora* CCA, CCB, CCP, CR1 e CR2. Havendo discrepâncias que evidenciem desvios significativos, isso poderá ser considerado um fator que contribui para a dificuldade de compreensão dos textos do CE. Após, com base nessas constatações, apresentaremos alternativas de reescrita para essas Sentenças se tornarem mais próximas dos padrões lexicais do PPE.

A seguir, na Figura 5.1, os *corpora* foram organizados numa linha horizontal e colorida em tons gradientes. A linha começa com os *corpora* que representam o PPE, o CCB e o CCP. À medida que os tons vão escurecendo, vão representando os *corpora* cuja linguagem vai se aproximando de um português mais erudito e menos atual. A ideia é verificar para qual lado dessa linha pende o nosso CE e, caso seja necessário, propor alternativas para aproximá-lo de um padrão semelhante aos *corpora* representantes do PPE.

Figura 5.1 – Corpora de Pesquisa



Fonte: Produção da autora³⁸

Trazemos como exemplo o emprego da expressão em negrito no trecho a seguir, extraído de uma Sentença judicial da Justiça Eleitoral do Brasil:

Contestação de Fulano de Tal, às fls. 220/230, arguindo a preliminar de inépcia da inicial, **ao aspecto** de ser omissa e contraditória, não constando a narração circunstanciada das supostas infrações imputadas ao contestante. (Recurso Eleitoral nº 0000011-34.2017.6.19.0151 – RIO DE JANEIRO, 2018, grifos nossos)

A expressão **ao aspecto de** está sendo empregada com o significado de *em razão de, em função de, devido a, por, etc.*, que expressam uma ideia de causa, de razão. Procuramos por essa expressão com este significado no CR2 e não a encontramos. Procuramos no CR1 e encontramos 4 ocorrências, duas em obras literárias de 1859 e duas em obras de 1891. Ou seja, neste caso, o texto se aproxima muito de um português do séc. XIX.

Como tentativa de aproximar o referido trecho a um padrão de português mais atual, sugerimos a seguinte reescrita: “Às fls. 220/230, Fulano de Tal, citado, contestou dizendo, preliminarmente, que o pedido inicial é omissa e contraditório e, por isso, não é válido. Alega que não há narração detalhada das supostas infrações atribuídas ao contestante”.

³⁸ Todas as figuras, as tabelas, os gráficos e os quadros cujas fontes não forem indicadas foram produzidas pela autora.

Expostos brevemente os *corpora* que farão parte de nossas análises e as razões pelas quais os escolhemos, passaremos, no próximo item, a detalhar os procedimentos metodológicos empregados. A partir daqui, nossa exposição se dará em subitens a seguir discriminados: no item 5.1, descreveremos os nossos *corpora*; no item 5.2, explicaremos como esses *corpora* são constituídos; no item 5.3, descreveremos as ferramentas utilizadas na análise.

5.1 DESCRIÇÃO DOS *CORPORA*

Como mencionamos no início desta seção, os nossos *corpora* se dividem em seis: *corpus* de Sentenças dos JECs do PJRS (CE); *corpus* de jornais tradicionais – Zero Hora e Folha de São Paulo (CCA) –, *corpus* de jornais populares – Diário Gaúcho e Massa (CCB) –, CorPop (CCP), *Corpus* de Referência 1 (CR1) e *Corpus* de Referência 2 (CR2).

A seguir, nos itens 5.1.1, 5.1.2 e 5.1.3, passaremos à descrição de cada um deles.

5.1.1 *Corpus* de Estudo - CE

O *corpus* de estudo (doravante CE) está constituído de Sentenças judiciais dos JECs do PJRS. Essas Sentenças representam as decisões tomadas pelos juízes em relação aos pedidos das pessoas que procuram os JECs para resolver algum conflito.

Conforme o art. 203 do Novo Código de Processo Civil (BRASIL, 2002), a Sentença judicial constitui um dos pronunciamentos do juiz e, segundo o art. 489, incisos I, II e III, desse mesmo Código, deve ser composta pelos seguintes elementos: o relatório, a fundamentação e o dispositivo.

No relatório, deve constar o nome das partes, dados de identificação do caso, o resumo do pedido da pessoa que entrou com a ação na Justiça e o resumo da contestação da outra parte envolvida no caso, além do registro das principais ocorrências havidas no andamento do processo. Na fundamentação, deve constar a análise do juiz quanto aos pedidos das partes envolvidas na ação e as razões que o levaram a tomar a decisão expressa no dispositivo.

No que se refere às Sentenças dos Juizados Especiais, porém, há algumas particularidades. Uma delas é a possibilidade de nelas não precisar constar o relatório. Segundo o art. 38 da Lei n. 9.099/95 (BRASIL, 1995), não há necessidade do relatório, ainda que muitas vezes seja necessário um breve relato dos fatos relacionados ao processo. Porém, os demais elementos – fundamentação e dispositivo – são indispensáveis.

A outra particularidade refere-se à possibilidade de a sentença ser proferida por um juiz leigo, e não um juiz togado. O juiz togado é aquele que ingressou na carreira por meio de concurso público, e o juiz leigo, segundo o art. 7º da Lei n. 9.099/95 (BRASIL, 1995), é considerado um auxiliar da Justiça e deve ser um advogado com mais de cinco anos de experiência. A grande maioria das Sentenças dos JECs do PJRS são proferidas por esses juízes leigos.

A seguir, na Figura 5.2, trazemos um exemplo de sentença desses Juizados Especiais Cíveis, com a sinalização de sua estrutura.

<p>COMARCA DE PORTO ALEGRE Posto Escola AJURIS Adjunto ao 5º Juizado Especial Cível Rua Celeste Gobato, 229, térreo – CEP: 90110160 Fone: 51-3284-9038</p> <hr/> <p>Processo n.: XXXXXXXX Autor: XXXXX Ré: XXXXX. Juíza Leiga: XXXXXX Data: 03 de agosto de 2015</p> <hr/> <p>Vistos etc.</p> <p><u>I - RELATÓRIO</u></p> <p>Relatório sucinto em face da previsão do artigo 38 da Lei 9.099/95.</p> <p>Trata-se de ação desconstitutiva de débito, com pedido liminar, objetivando o autor, LEONARDO SALERMO, a declaração de inexistência e ilicitude dos débitos nos valores de R\$ 33,92 (trinta e três reais e noventa e dois centavos) e R\$ 3,17 (três reais e dezessete centavos), apontados pela ré no CPF do autor como devidos e não pagos, bem como a indenização por danos morais, em face da ré, GLOBAL VILLAGE TELECOM LTDA.</p> <p>É o relatório.</p> <p><u>II- FUNDAMENTAÇÃO</u></p> <p><u>2.1 - Da relação de consumo e da responsabilidade</u></p> <p>As partes se enquadram nos conceitos de consumidor e fornecedor de acordo com as descrições previstas, respectivamente, nos arts. 2º e 3º CDC. [...]</p> <p><u>2.2 - Do defeito no serviço</u></p> <p>A lide versa a respeito de defeito do serviço, de acordo com art. 14, “caput”, do Código de Defesa do Consumidor, [...]</p>	<p>Nesse sentido é a jurisprudência das Turmas Recursais Cíveis do Rio Grande do Sul:</p> <p>RECURSO INOMINADO. TELEFONIA. AÇÃO DE DESCONSTITUIÇÃO DE DÉBITO C/C INDENIZAÇÃO POR DANOS MORAIS. ALEGAÇÃO DE COBRANÇA INDEVIDA. ÔNUS DA DEMANDADA, ART. 333, II, CPC. AUSENTE DEMONSTRAÇÃO DA REGULARIDADE DOS DÉBITOS. COBRANÇA INDEVIDA CONFIGURADA. INSCRIÇÃO INDEVIDA. DANO MORAL MANTIDO. Análise das faturas de fl. 17/28 que corroboram a alegação da parte autora, pois mostram cobranças concomitantes da empresa ora recorrente e GVT, em virtude de frustrada tentativa de portabilidade. Recurso da requerida que se mostra genérico. Defesa superficial da regularidade dos débitos. Cobrança indevida. [...]</p> <p>ISTO POSTO, opino pela PROCEDÊNCIA do pedido da parte autora LEONARDO SALERMO frente à parte ré GLOBAL VILLAGE TELECOM LTDA. a) para desconstituir todos os débitos da contratação referida; b) para condenar ao pagamento de R\$ 7.880,00 (sete mil oitocentos e oitenta reais), por dano extrapatrimonial, incidindo correção monetária pelo índice do IGP-M a partir da sentença e juros moratórios de 1% (um por cento) ao mês a partir da citação; c) para determinar a expedição de ofício a SERASA para excluir o nome da parte autora, junto ao seu cadastro, referente ao negócio jurídico em questão.</p> <p>Nos termos do que dispõe o artigo 55 da Lei dos Juizados Especiais Cíveis (Lei nº 9.099/95), não há condenação da parte sucumbente ao pagamento de custas processuais e honorários de advogado.</p> <p>Porto Alegre, 03 de agosto de 2015.</p> <p>XXXXXXXXXXXX, Juíza Leiga</p>
--	--

Em vermelho, está o cabeçalho, com os dados de identificação (número do processo, nome das partes, local, nome do juiz, etc.). Em seguida, está a expressão *Vistos etc.*, que serve para indicar que os atos anteriores à sentença foram devidamente analisados. Após, em azul, aparece o relatório, que, como já dissemos, é dispensável e muitas vezes nem consta nessas Sentenças.

Em marrom, vem a parte da fundamentação, e nela sinalizamos um trecho roxo. Trata-se de uma ementa jurisprudencial referente ao mesmo tema abordado na sentença. A ementa jurisprudencial traz o resumo de uma decisão judicial anterior e é muito usada para fundamentar as razões do juiz para decidir. Normalmente, ela é composta de um cabeçalho e de um dispositivo. No cabeçalho, frequentemente em letras maiúsculas, há apenas frases nominais – conjunto de palavras-chave – que indicam o tipo de recurso e o assunto. No dispositivo, com frases verbais, se apresenta a tese jurídica que fundamenta a decisão.

Por fim, em preto, se encontra o dispositivo, ou seja, a parte destinada a expor a decisão do juiz, seguida da data em que foi proferida essa decisão e o nome do juiz, com a qualificação de Juiz Leigo ou Juiz Togado.

Cumpra também referir alguns aspectos relativos à microestrutura, como o fecho do relatório *É o relatório*, que evidencia que ali se encerra o relato dos fatos relacionados ao processo. Na parte da fundamentação, nesta sentença, há várias subdivisões de acordo com o assunto a ser tratado, como se vê, por exemplo em *2.1 – Da relação de consumo e da responsabilidade*. Ainda na fundamentação, a frase *Nesse sentido é jurisprudência das Turmas Recursais Cíveis do Rio Grande do Sul* é típica para introduzir citações de ementas jurisprudenciais. Além disso, no início da parte dispositiva, há a expressão *ISTO POSTO*, normalmente empregada nas Sentenças para introduzir a conclusão do texto com a decisão da sentença.

5.1.2 Corpus de Contraste CCA

O *corpus* de contraste CCA é constituído de textos de dois jornais tradicionais: Zero Hora e Folha de São Paulo. Ambos são considerados jornais de referência, por serem consolidados e com grande circulação no país, além de serem dirigidos a classes de maior poder aquisitivo, quais sejam, classes A, B e C. Apresentam uma linguagem segundo a norma culta da língua.

Cabe referir que, segundo a Wikipedia, o jornal Zero Hora

é um dos maiores jornais de circulação diária do Brasil. É editado em Porto Alegre e controlado pelo Grupo RBS. Ao lado de *Folha de S. Paulo*, *Estado de Minas*, *O Globo*, *Correio Braziliense* e *O Estado de S. Paulo*, entre outros, forma o grupo dos principais jornais de referência do Brasil.³⁹

E o jornal Folha de São Paulo,

também conhecida como Folha de São Paulo ou simplesmente Folha, é um jornal brasileiro editado na cidade de São Paulo e jornal de maior circulação do Brasil, com 340.511 exemplares (incluindo assinantes digitais), segundo o Instituto Verificador de Comunicação (IVC), em junho de 2020.⁴⁰

A seguir, a título de ilustração, colocamos as Figuras 5.3 e 5.4, com as imagens desses jornais na versão impressa disponível *on-line* para assinantes.

Figura 5.3 – Zero Hora



Fonte: Site do jornal Zero Hora – 20-03-2017

Figura 5.4 – Folha de São Paulo



Fonte: Site do jornal Folha de São Paulo, 20-03-2017

5.1.3 Corpus de Contraste CCB

O *corpus* de contraste CCB está constituído de textos de dois jornais populares: Diário Gaúcho e Massa. O primeiro, Diário Gaúcho (DG), é publicado pela empresa jornalística RBS, em Porto Alegre, com versão impressa e *on-line*. O segundo, Massa

³⁹ **Zero Hora**. In: WIKIPÉDIA, a enciclopédia livre. Flórida: Wikimedia Foundation, 2022. Disponível em: https://pt.wikipedia.org/w/index.php?title=Zero_Hora&oldid=63418976. Acesso em: 07 maio 2022.

⁴⁰ **Folha de São Paulo**. In: WIKIPÉDIA, a enciclopédia livre. Flórida: Wikimedia Foundation, 2022. Disponível em: https://pt.wikipedia.org/wiki/Folha_de_S.Paulo. Acesso em: 07 maio 2022.

(JM), publicado na Bahia pelo grupo A TARDE, em versão impressa e *on-line*. Ambos são considerados jornais populares por terem por objetivo oferecer notícias e informação de entretenimento e serviço às classes C, D e E. Para tanto, valem-se de muitas imagens, textos curtos com uma linguagem em princípio mais simples que as dos jornais do CCA.

A seguir, a título de ilustração, colocamos, nas Figuras 5.5 e 5.6, as imagens de capas desses jornais na versão impressa disponível *on-line*.

Figura 5.5 – Diário Gaúcho



Fonte: Site do Diário Gaúcho – 21-03-2017

Figura 5.6 – Jornal Massa



Fonte: Site do Jornal Massa – 21-03-2017

Pelas imagens das Figuras 5.3, 5.4, 5.5 e 5.6, é possível perceber as diferenças na linguagem empregada de acordo com a proposta de cada jornal. No Jornal Zero Hora (Figura 5.3), sublinhamos a expressão *planta frigorífica*, que nos parece técnica; no Jornal Folha de São Paulo (Figura 5.4), tem-se o verbo *minimiza*, que também sublinhamos porque não nos parece ser empregado pelo cidadão comum brasileiro, com pouca escolaridade e pouca experiência em leitura. Encontra-se no CCP e no CR2, mas não no CR1. No Jornal Massa (Figura 5.6), chama a atenção o emprego de expressões bem populares como *galera*, *tá na bruxa*, *não larga o osso*, *níver*, o que parece ir ao encontro do objetivo do jornal, qual seja, o de atingir o povo com uma linguagem escrita muito próxima da empregada na fala. Porém, no corpo das notícias a linguagem empregada é mais formal, sem usar esses recursos apelativos.

5.1.4 *Corpus de Contraste CCP*

O *Corpus de Contraste CCP - CorPop* (PASQUALINI, 2018) - foi compilado com base no nível de letramento médio dos leitores do país e das características que poderiam compor um padrão de simplicidade textual em um *corpus* de textos adequados a esses leitores. Essas características foram levantadas com bases nas pesquisas do INAF, nos resultados de pesquisas relativas ao Retratos da Leitura no Brasil e em um questionário com leitores dentro do perfil sociodemográfico condizente com as características do *corpus*.

Está composto por:

- (1) textos do jornalismo popular (DG e Jornal Hora de Santa Catarina);
- (2) textos e autores mais lidos pelos respondentes das últimas edições da pesquisa Retratos da Leitura no Brasil;
- (3) textos da coleção “É Só o Começo”, que envolve adaptação de clássicos da literatura brasileira para leitores com baixo letramento, adaptação produzida por linguistas;
- (4) textos do jornal *Boca de Rua*, produzido por pessoas em situação de rua, com baixa escolaridade e baixo letramento sob a supervisão de uma jornalista; e,
- (5) textos do *Diário da Causa Operária*, veículo da imprensa operária brasileira, um jornal produzido também por pessoas dentro da faixa média de letramento dos leitores do nosso país.

Por se constituir desses materiais-fonte diferenciados, o CorPop⁴¹ serve para oferecer uma referência para o português popular brasileiro escrito, que é igualmente culto, em uso pela maioria dos brasileiros, e que extrapola o gênero jornalístico.

Na Tabela 5.1, abaixo, encontra-se o conteúdo do CorPop em módulos e a quantidade de *types*⁴² e *tokens*⁴³.

⁴¹ As listas de palavras – brutas e lematizadas – desse material bem como o acesso a esse material mediante expressões de busca são possíveis em <http://www.ufrgs.br/texteccc/porlexbras/corpop/ferramentas.php>.

⁴² *Types* refere-se a palavras diferentes do texto.

⁴³ *Tokens* refere-se ao total de palavras do texto.

Tabela 5.1 – Conteúdo do CorPop

Módulo	Types Palavras diferentes	Tokens Total de palavras - formas
PorPopular Diário Gaúcho	6.378	30.944
Jornal Hora de Santa Catarina	4.118	18.303
Jornal Boca de Rua	8.913	71.454
Jornal Diário da Causa Operária	7.841	59.785
Textos de estudo Retratos da Leitura no Brasil	22.463	430.806
Textos da Coleção "É Só o Começo"	8.161	73.507
Total	32.138	684.799

Fonte: <http://www.ufrgs.br/textecc/porlexbras/corpop/>. Acesso em: 05 out. 2018.

Aplicando-se ao CorPop à caracterização dos *corpora* proposta por Berber Sardinha e exemplificada no Quadro 4.1 deste trabalho, temos a configuração proposta no Quadro 5.1

Quadro 5.1 – Caracterização do CorPop (CCP)

MODO	Escrito	
TEMPO	Sincrônico – Contemporâneo	
SELEÇÃO	De amostragem/Estático	Amostra finita.
CONTEÚDO	Língua Geral	Os textos representam o português brasileiro popular escrito.
	Regional ou dialetal	
AUTORIA	De língua nativa	Os autores são falantes nativos.
	Mais de um autor	O texto é produzido por mais de um indivíduo.
	Institucional	Os textos dos jornais são produzidos por empresas jornalísticas.
FINALIDADE	De contraste	O <i>corpus</i> é usado para fins de contraste com o <i>corpus</i> de estudo.

Fonte: Produzido pela autora com base em BERBER SARDINHA (2004, p. 20-2); DUARTE (2008, p.161-3).

5.1.5 *Corpus* de Referência 1 – CR1

O *Corpus* de Referência 1 (CR1) é composto por obras literárias brasileiras que estão em domínio público⁴⁴. Por serem obras publicadas entre os anos de 1840 e 1956, apresentam uma linguagem clássica mais antiga. O *corpus* contém 2.416.960 palavras, segundo a contagem do editor de texto *Word*. E, segundo o AntConc 3.2.1w⁴⁵, 2.460.268 *tokens* e 82.136 *types*.

No Quadro 5.2, abaixo, está a lista de todos os autores e obras deste *corpus*.

Quadro 5.2 – Conteúdo do *Corpus* de Referência 1 (CR1)

Autor	Obras	Ano de Publicação
Aluísio Azevedo	A Mortalha de Alzira	1894
	O Cortiço	1890
	O Mulato	1881
	Uma Lágrima de Mulher	1879
Euclides da Cunha	Contrastes e Confrontos	1907
	Os Sertões - Vol. 1	1902
Franklin Távora	O Sacrifício	1879
	O Matuto	1902
	O Cabeleira	1876
Inglês de Souza	O Missionário	1899
Joaquim Macedo	As Vítimas Algozes	1869
	As Mulheres de Mantilha	1870
	Amor e Pátria	1859
	A Moreninha	1844
	Luneta Mágica	1869
José de Alencar	A Alma de Lázaro	1848
	A Viuvinha	1857
	Iracema	1865
	O Guarani	1857
	Senhora	1875
Lima Barreto	A Nova Califórnia	1915
	Clara dos Anjos	1948

⁴⁴ As obras constituintes do CR1 foram extraídas do site Domínio Público, disponível em www.dominiopublico.gov.br. Este *site*, criado em 2004, disponibiliza para o público em geral uma biblioteca virtual, com obras literárias, artísticas e científicas (na forma de textos, sons, imagens e vídeos).

⁴⁵ Trata-se de um *software* livre que fornece listas de palavras, concordâncias e outras utilidades para pesquisas em *corpora* linguísticos. Disponível em <http://www.laurenceanthony.net/software.html>

	O Triste Fim de Policarpo Quaresma	1915
	Recordações do Escrivão Isaías Caminha	1909
	Vida Urbana	1956
Machado de Assis	Ressurreição	1872
	Helena	1876
	Iaiá Garcia	1878
	Memórias Póstumas de Brás Cubas	1881
	Quincas Borba	1886
	Memorial de Aires	1908
	Dom Casmurro	1899
	A mão e a luva	1874
Paulo Setúbal	A Marquesa dos Santos	1925
	Alma Cabocla	1920
	Maluquices do Imperador	1927
	O Romance do Prata	1935
	O Sonho das Esmeraldas	1935
	O Príncipe de Nassau	1926
	Os Irmãos Leme	1933
Raul Pompeia	As Jóias da Coroa	1882
	O Ateneu	1888
	Uma Tragédia no Amazonas	1880
Visconde de Taunay	A retirada de Laguna	1874
	Ao Entardecer	1901
	Inocência	1872

Aplicando-se ao CR1 a caracterização dos *corpora* exemplificada no Quadro 4.1 deste trabalho, teremos a configuração proposta no Quadro 5.3.

Quadro 5.3 – Caracterização do CR1

MODO	Escrito	
TEMPO	Diacrônico	
SELEÇÃO	De amostragem/Estático	Amostra finita.
CONTEÚDO	Língua Geral	Os textos representam o português brasileiro do gênero literário.
	Regional ou dialetal	
AUTORIA	De língua nativa	Os autores são falantes nativos.
	Mais de um autor	O texto é produzido por mais de um indivíduo.
FINALIDADE	De referência	O <i>corpus</i> , no mínimo cinco vezes maior que o <i>corpus</i> de estudo, é usado para fins

		de contraste com o <i>corpus</i> de estudo.
--	--	---

Fonte: Produzido pela autora com base em BERBER SARDINHA (2004, p. 20-2); DUARTE (2008, p.161-3).

5.1.6 *Corpus* de Referência 2 – CR2

O *Corpus* ptTenTen: *Corpus of the Portuguese Web* (CR2) é composto por um conjunto de textos coletados da internet. São textos do Português Brasileiro (76%) e do Português Europeu (24%)⁴⁶. É um *corpus* construído com textos da web, coletados entre março de 2011 e agosto de 2012, por meio do mesmo método e utilizando tecnologia especializada em coletar apenas conteúdo *web* linguisticamente valioso. Ou seja, são descartados conteúdos duplicados, frases incompletas, anúncios, etc., que não sejam pertinentes ao texto em foco.

Está disponível para consulta no sistema *Sketch Engine*⁴⁷, uma plataforma de *software* comercial que fornece lista de palavras por critérios variados – frequência, classe gramatical etc. –, *thesaurus* de distribuição, concordâncias, extração de terminologias, criação e comparação de *corpora*, entre outros.

Segundo informações no *Sketch Engine*, o CR2 apresenta 4.622.750.491 *tokens*⁴⁸ e 3.896.392.719 palavras⁴⁹.

Aplicando-se ao CR2 a caracterização dos *corpora* exemplificada no Quadro 4.1 deste trabalho, teremos a configuração proposta no Quadro 5.4.

Quadro 5.4 – Caracterização do CR2

MODO	Escrito	
TEMPO	Sincrônico – Contemporâneo	
SELEÇÃO	De amostragem/Estático	Amostra finita.
CONTEÚDO	Língua Geral	Os textos representam o português do Brasil e de Portugal.
	Regional ou dialetal	
AUTORIA	De língua nativa	Os autores são falantes nativos.

⁴⁶ Apesar de o PtTenTen conter 24% de textos do português europeu, nós o escolhemos porque seus textos foram coletados da internet. Além disso, os 76% de textos do português brasileiro representam aproximadamente 3 bilhões de palavras.

⁴⁷ **Sketch Engine**. Disponível em <https://auth.sketchengine.eu/#login>. Acesso em: 15 jun. 2020.

⁴⁸ O Sketch Engine considera *tokens* não apenas as palavras, mas também outros caracteres como sinais de pontuação, etc.

⁴⁹ Aqui a contagem do Sketch Engine é apenas em relação às palavras do documento.

	Autores variados	Os textos podem ter autores únicos como múltiplos
	Institucional	Os textos provêm de variadas fontes da web.
FINALIDADE	De referência	O <i>corpus</i> , no mínimo cinco vezes maior que o <i>corpus</i> de estudo, é usado para fins de contraste com o <i>corpus</i> de estudo.
CODIFICAÇÃO		O texto apresenta informações extratextuais como cabeçalho e etiquetas (morfo sintáticas, sintáticas, semânticas, discursivas ou pragmáticas)

Fonte: Produzido pela autora com base em BERBER SARDINHA (2004, p. 20-2); DUARTE (2008, p.161-3).

A seguir, detalharemos como se deu a compilação dos *corpora* CE, CCA, CCB e CR1.

5.2 COMPILAÇÃO DOS *CORPORA* CE, CCA, CCB E CR1

Nesta seção apresentamos a forma como compilamos os *corpora* CE, CCA, CCB e CR1. No item 5.2.1, descrevemos a compilação do nosso CE; no item 5.2.2, o *corpus* CCA; no item 5.2.3, o *corpus* CCB; e, no item 5.2.4, o *corpus* CR1.

5.2.1 *Corpus* de Estudo

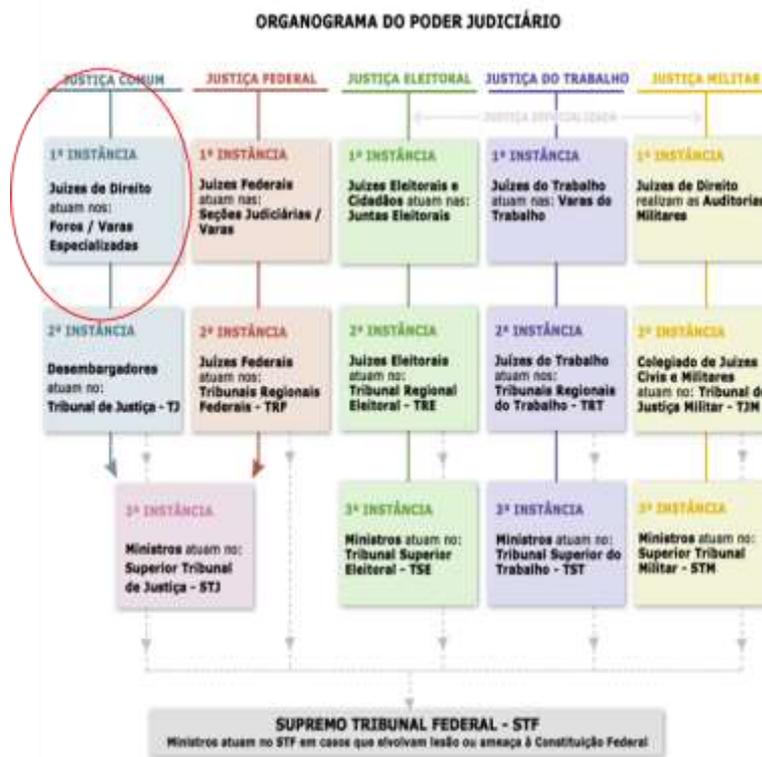
As Sentenças que constituem o nosso CE foram coletadas do *site* do Tribunal de Justiça do Rio Grande do Sul (PJRS). Estando na *internet*, tais textos são de domínio público. Além disso, dispõe o inciso IX do artigo 93 da Constituição de 1988:

Todos os julgamentos dos órgãos do Poder Judiciário serão públicos, e fundamentadas todas as decisões, sob pena de nulidade, podendo a lei limitar a presença, em determinados atos, às próprias partes e a seus advogados, ou somente a estes, em casos nos quais a preservação do direito à intimidade do interessado no sigilo não prejudique o interesse público à informação. (BRASIL. Constituição, 1988).

Antes de prosseguirmos no detalhamento da coleta do nosso CE, vamos primeiro delimitar o âmbito de nossa pesquisa: Juizados Especiais Cíveis do Poder Judiciário do

Rio Grande do Sul. Para tanto, colocamos a seguir a Figura 5.7, em que consta o organograma do Poder Judiciário no Brasil.

Figura 5.7 – Organograma do Poder Judiciário no Brasil



Fonte: <http://nev.incubadora.fapesp.br/porta/segurancajustica/judiciario/organogramadojudiciario> – Acesso em: 25 set. 2009

De acordo com o organograma acima, verificamos que o Poder Judiciário brasileiro é composto por vários órgãos, cingindo-se o nosso trabalho ao âmbito da Justiça Comum, marcada por uma elipse na figura acima. A Justiça Comum é constituída pelos Tribunais de Justiça e pelos Juízes de Direito e tem competência residual. Ou seja, todas as matérias que não sejam de competência específica de outros órgãos jurisdicionais (Justiça Federal, Justiça do Trabalho e Justiça Eleitoral) competem à Justiça Comum.

Integram ainda a Justiça Estadual os Juizados Especiais Cíveis e Criminais. O Juizado Especial Cível tem como atribuição a conciliação, o processo e o julgamento de ações cíveis de menor complexidade, cujo valor de causa não ultrapasse 40 salários-mínimos. O Juizado Especial Criminal se encarrega de conciliar e julgar as infrações penais de menor potencial ofensivo, ou seja, aquelas em que a pena máxima não ultrapasse dois anos. A instância superior dos Juizados Especiais são as Turmas Recursais, que, apesar de órgão recursal, não são consideradas integrantes do Tribunal de Justiça.

O Tribunal de Justiça do Estado do Rio Grande do Sul representa a segunda instância da Justiça Comum. Nele são julgados os recursos provenientes da Justiça de primeiro grau, ou primeira instância, que, em regra, é a porta de entrada de todas as demandas judiciais. A partir da decisão de primeiro grau, concretizada na sentença do juiz, a parte que se sente prejudicada tem a possibilidade de recorrer a uma instância superior, que, no caso, é o Tribunal de Justiça.

Na medida em que centramos a coleta de nossas Sentenças nos Juizados Especiais Cíveis, cumpre-nos também esclarecer que, no que se refere à ciência do Direito e suas ramificações, o Direito,

[...] em seu sentido objetivo [...] apresenta-se como um complexo orgânico, cujo conteúdo é constituído pela soma de preceitos, regras e leis, com as respectivas sanções, que regem as relações do homem, vivendo em sociedade [...] E se diz Direito Público ou Direito Privado, Direito Constitucional, Direito Administrativo, Direito Penal, Direito Processual, Direito Civil, Direito Comercial. (SILVA, 2007, eletrônico).

Percebemos, por esta citação, que o Direito se ramifica em várias áreas, de acordo com a matéria de que trata. Em sentido amplo, é possível afirmar que o Direito se divide em dois grandes ramos: Direito Público e Direito Privado.

Pelo exposto, por serem nossas Sentenças provenientes dos Juizados Especiais Cíveis do Poder Judiciário do Rio Grande do Sul, tratam de matérias relativas ao Direito Civil. Assim, aportam a esses Juizados uma variedade enorme de pedidos relativos a conflitos de menor complexidade: conflitos que não dependem de vistorias, ou perícias, ou outro tipo de prova mais elaborada, além de terem 40 salários mínimos como valor máximo da causa. Vale ressaltar que os Juizados Especiais Cíveis, conforme o § 2º do art. 3º da Lei 9.099/95, não podem atuar em muitas ações, como, por exemplo, ações de alimentos (pagamento de pensão alimentícia), separação e divórcio, ações que envolvam problemas da Infância e Juventude, questões de recuperação judicial de empresas, inventários etc.

Para acessar as Sentenças dos JECs do PJRS, é necessário ter-se o número do processo, ou o número da OAB do advogado, ou o nome de uma das partes envolvidas. Como não dispúnhamos desses dados, foram objeto de coleta para composição do nosso CE Sentenças dos JECs que foram sujeitas a recurso. Isto é, compõem o nosso CE Sentenças cuja decisão não satisfiz a uma ou a ambas as partes e que, por isso, foram submetidas a reexame pelas Turmas Recursais Cíveis, grau de recurso desses Juizados. Sendo submetidas a reexame pelas Turmas Recursais, é possível, via *site* do PJRS, na

página de Pesquisa de Jurisprudência⁵⁰, acessar essas Sentenças sem a necessidade dos dados anteriores.

Conforme já mencionamos, as Turmas Recursais representam a instância superior dos Juizados Especiais. Elas se dividem, de acordo com a temática de que tratam, em Turmas Recursais Cíveis, Turmas Recursais Criminais e Turmas Recursais da Fazenda Pública. Assim, como as Sentenças objeto do nosso estudo são as dos Juizados Especiais Cíveis, foram as Turmas Recursais Cíveis que serviram como porta de acesso às Sentenças do nosso CE. No PJRS, há quatro Turmas Recursais Cíveis: Primeira Turma Recursal Cível (doravante 1TRC); Segunda Turma Recursal Cível (doravante 2TRC); Terceira Turma Recursal Cível (doravante 3TRC) e Quarta Turma Recursal Cível (doravante 4TRC).

Estabelecida a via de acesso às Sentenças do nosso CE, estipulamos como período de coleta os anos de 2018 e 2019, dividindo cada um desses períodos anuais em 11 períodos mensais, a partir do início dos trabalhos forenses, que se dá no dia 20 de janeiro, até o início do recesso forense, que se dá no dia 20 de dezembro. Assim, nossos períodos anuais de pesquisa ficaram divididos em subperíodos como segue: 20-01 a 19-02; 20-02 a 19-03; até chegar a 20-12.

Há que esclarecer também que as datas das Sentenças coletadas são sempre anteriores às do período de coleta. Isso se deve ao fato de o período de coleta ser relativo ao julgamento do recurso referente à sentença e não à data em que esta foi proferida. Diante disso, poderá haver Sentenças proferidas no ano de 2015, por exemplo, correspondendo a três anos de antecedência em relação ao ano de 2018, um dos anos da coleta.

Considerando, então, as quatro Turmas Recursais Cíveis que serviram como porta de acesso às nossas Sentenças e os subperíodos anuais que discriminamos acima, optamos por selecionar 5 Sentenças de cada uma das Turmas Recursais em cada subperíodo. Assim, num período de dois anos, ficamos com 440 Sentenças.

A seguir, colocamos a Tabela 5.2, que ilustra os procedimentos explanados acima.

⁵⁰ Jurisprudência significa, entre outras coisas, o conjunto das decisões dos tribunais. No caso, a jurisprudência é a relativa às decisões das Turmas Recursais Cíveis.

Tabela 5.2 – Sentenças do CE por Período

PERÍODOS	ANOS	TURMAS RECURSAIS CÍVEIS				
		1TRC	2TRC	3TRC S	4TRC	TOTAL DE SENTENÇAS
20-01 a 19-02	2018	5	5	5	5	20
	2019	5	5	5	5	20
20-02 a 19-03	2018	5	5	5	5	20
	2019	5	5	5	5	20
20-02 a 19-03	2018	5	5	5	5	20
	2019	5	5	5	5	20
20-04 a 19-05	2018	5	5	5	5	20
	2019	5	5	5	5	20
20-05 a 19-06	2018	5	5	5	5	20
	2019	5	5	5	5	20
20-06 a 19-07	2018	5	5	5	5	20
	2019	5	5	5	5	20
20-07 a 19-08	2018	5	5	5	5	20
	2019	5	5	5	5	20
20-08 a 19-09	2018	5	5	5	5	20
	2019	5	5	5	5	20
20-09 a 19-10	2018	5	5	5	5	20
	2019	5	5	5	5	20
20-10 a 19-11	2018	5	5	5	5	20
	2019	5	5	5	5	20
20-11 a 19-12	2018	5	5	5	5	20
	2019	5	5	5	5	20
TOTAL DE SENTENÇAS		110	110	110	110	440

Na Tabela 5.3, apresentamos dados relativos ao número de palavras dos conjuntos de sentenças do nosso CE. Tais dados são referentes à contagem de palavras do editor de texto *Word* e se encontram discriminados por Turma Recursal e por período anual. Na última coluna da tabela, encontra-se o número total de palavras cada Turma, e, na última linha, encontra-se o número total de palavras por ano. O CE ficou com um total de

547.602⁵¹ palavras segundo a contagem do editor de texto Word. Submetido à ferramenta AntConc, ficou com 542.836 *tokens* e 14.665 *types*.

Tabela 5.3 – Número de palavras do CE por Turma Recursal e por Período

TURMA	2018	2019	TOTAL
1TRC	63.895	71.253	135.148
2TRC	75.068	64.422	139.490
3TRC	73.513	62.168	135.681
4TRC	74.086	65.038	139.124
TOTAL	286.562	262.881	549.443

Outro critério que utilizamos para a seleção das Sentenças foi a natureza da demanda. Como já referimos neste trabalho, os JECs tratam de matérias relativas ao Direito Civil, muitas delas referentes à Responsabilidade Civil. A temática da Responsabilidade Civil é muito ampla e está presente em praticamente todas as áreas do Direito, quer Público, quer Privado. E isso pode ser conferido, por exemplo, no Relatório Anual do Tribunal de Justiça do Rio Grande do Sul de 2017⁵², em que esse tema figura em quarto lugar entre os processos que mais foram julgados pelo TJRS em 2017.

Ainda que tal relatório não discrimine a natureza da demanda por órgão julgador, as demandas que chegam aos JECs são, em sua maioria, relativas a cadastro de inadimplentes, prestação de serviços em geral, energia elétrica, telefonia, seguros etc. São matérias que se inserem no âmbito da Responsabilidade Civil.

Assim, para cada grupo de 5 Sentenças relativas a cada subperíodo e a cada Turma Recursal, conforme a Tabela 5.2, selecionamos demandas que envolvessem as temáticas a seguir:

- a) **acidente de trânsito** – problemas relativos a acidentes de trânsito;
- b) **água** – problemas relativos a fornecimento e serviços de água para o cidadão;
- c) **cadastro de inadimplentes** – problemas relativos à inscrição negativa do cidadão em órgãos de proteção ao crédito;
- d) **energia elétrica** – problemas relativos a fornecimento e serviços de energia elétrica para o cidadão;
- e) **plano de saúde** – problemas envolvendo planos de saúde, como negativa de

⁵¹ O número de 547.602 palavras foi a leitura realizada pelo editor de texto *Word* em um arquivo único com todas as sentenças do CE. Ele não confere com o número total de palavras da Tabela 5.3, que ficou em 549.443. Como essa contagem é feita por ferramentas computacionais, alguma alteração na formatação pode gerar diferenças como a que ocorreu aqui.

⁵² Disponível em:

https://www.tjrs.jus.br/site/administracao/prestacao_de_contas/relatorio_anual/2017/pdf/RA_2017-9-Relatorios_Estatisticos_v2.pdf. Acesso em: 17 nov. 2018.

- cobertura para algum procedimento, aumento abusivo de mensalidades, etc.;
- f) **prestação de serviços** – problemas relativos à prestação de serviços em geral, como o caso de não conclusão de uma obra, de um serviço malfeito, etc.;
- g) **telefonia** – problemas relativos a fornecimento e serviços de telefonia fixa e móvel para o cidadão.

Na Tabela 5.4, a seguir, ilustramos como ficaram as sentenças distribuídas por natureza da demanda.

Tabela 5.4 – Sentenças do CE por Temática

TIPO DE DEMANDA	NÚMERO DE SENTENÇAS
ÁGUA	45
CADASTRO DE INADIMPLENTES	55
ENERGIA ELÉTRICA	73
PLANO DE SAÚDE	68
PRESTAÇÃO DE SERVIÇOS	99
TELEFONIA	85
TOTAL DE SENTENÇAS	440

Acreditamos que, dessa forma, podemos contemplar uma maior variedade temática e, com isso, ampliar a representatividade do nosso CE.

A seguir, no Quadro 5.5, mostramos a classificação do nosso CE conforme os critérios utilizados por Berber Sardinha (2004) exemplificados no Quadro 4.1 deste trabalho.

Quadro 5.5 – Caracterização do CE

MODO	Escrito	
TEMPO	Sincrônico e Contemporâneo	
SELEÇÃO	De amostragem/Estático	Amostra finita.
	Equilibrado	Os componentes são distribuídos em quantidades semelhantes.
CONTEÚDO	Especializado	Os textos são de registros específicos: as sentenças judiciais constituem textos técnicos, com função normativa e informativa de especialista para leigo.
AUTORIA	De língua nativa	O art. 192 do Novo Código de Processo Civil preceitua: “Em todos os atos e termos do processo é obrigatório o uso da língua portuguesa” (BRASIL, 2002).

	Mais de um autor	As sentenças são, em sua maioria, redigidas pelo juiz leigo, mas homologadas pelo juiz togado.
	Institucional	O texto é produzido por uma instituição. As sentenças servem como documento informativo no âmbito do PJRS, sendo disponibilizadas no seu <i>site</i> .
FINALIDADE	De estudo	O <i>corpus</i> que se pretende descrever.

Fonte: Produzido pela autora com base em BERBER SARDINHA (2004, p. 20-2); DUARTE (2008, p.161-3).

E agora, para ilustrarmos o caminho percorrido para a coleta de nossas Sentenças, abaixo colocamos algumas figuras mostrando o passo a passo de nossa coleta.

Passo 1. Na página inicial do *site* do PJRS⁵³, clicamos em Pesquisa de Jurisprudência, tal como se vê na Figura 5.8, a seguir:

Figura 5.8 – Passo 1



Passo 2. Após, na página relativa à Pesquisa de Jurisprudência, tal como se vê na Figura 5.9, digitamos, no campo destinado à busca, a expressão Recurso Inominado. Usamos esta expressão porque a maioria dos recursos julgados nas Turmas Recursais recebe este nome. O recurso é qualificado de inominado, porque a Lei dos Juizados Especiais não atribui nome aos recursos que chegam às Turmas Recursais.

⁵³ Página inicial do *site* do PJRS. Disponível em: <https://www.tjrs.jus.br/site/>

Figura 5.9 – Passo 2



Passo 3. Em seguida, conforme se vê na Figura 5.10, no campo Tribunal (n. 1), clicamos em Turmas Recursais; no campo Órgão Julgador (n. 2), clicamos na Turma Recursal Cível que estávamos pesquisando; no campo Data do Julgamento (n. 3), colocamos o período desejado.

Figura 5.10 – Passo 3



Passo 4. Preenchidos esses campos e dando-se um *enter*, surge a próxima página, conforme Figura 5.11, a seguir. Nela vão aparecer as referências aos dez últimos recursos inominados julgados até aquela data.

Figura 5.11 – Passo 4



Passo 5. Clicando no primeiro número, tal como apontado pela seta, aparecerá outra janela, para inserção de um código, conforme Figura 5.12, a seguir.

Figura 5.12 – Passo 5



Passo 6. Surge, então, a janela a seguir, Figura 5.13. Nela, clicamos no número indicado pela seta. Este é o número do processo relativo à sentença.

Figura 5.13 – Passo 6



Passo 7. Em seguida, surge outra tela, conforme a Figura 5.14, a seguir.

Figura 5.14 – Passo 7



Clicamos onde consta *Ver Sentença*, conforme a seta está indicando. Aparece a tela da Figura 5.15, abaixo.

Figura 5.15 – Passo 8



Passo 8. Clicando no número indicado, temos acesso à sentença, cujo arquivo pode vir em formato .pdf ou .doc. Estando o arquivo em formato .pdf, alteramos para o formato .doc e em seguida salvamos em pastas no nosso computador.

Para o armazenamento de nossas Sentenças, reunidas como um *corpus* de estudo, adotamos os seguintes procedimentos:

- a. Inicialmente, criamos 4 pastas: 1TRC (1ª TURMA RECURSAL CÍVEL),

2TRC (2ª TURMA RECURSAL CÍVEL), 3TRC (3ª TURMA RECURSAL CÍVEL), 4TRC (4ª TURMA RECURSAL CÍVEL).

- b. Em cada uma dessas pastas, criamos duas subpastas. Na pasta 1TRC, por exemplo, criamos as pastas 1TRC-2018 (com as sentenças de 2018 da 1ª Turma Recursal Cível) e 1TRC-2019 (com as sentenças de 2019 da 1ª Turma Recursal Cível). E assim procedemos para as outras três pastas, referentes às outras Turmas Recursais Cíveis.
- c. Em cada uma dessas subpastas, criamos 2 pastas para cada período do ano coletado. Na subpasta denominada 1TRC-01-18, por exemplo, ficaram as sentenças originais do primeiro período de coleta da 1ª Turma Recursal Cível. Na subpasta denominada 1TRC-01-18-L, ficaram as sentenças limpas do primeiro período de coleta da 1ª Turma Recursal Cível. Assim fizemos para o ano de 2019 e para as demais Turmas Recursais.
- d. Para a limpeza das Sentenças, retiramos todos os dados de identificação do processo, como nome das partes, nome do julgador, número do processo. Também retiramos qualquer caractere ou formatação que possa impedir o processamento do texto pelas ferramentas que utilizaremos para as análises.
- e. As Sentenças foram salvas com o seguinte critério de denominação: S1-1TRC-01-18, em que S1 corresponde à Sentença 1; 1TRC, à 1ª Turma Recursal Cível; 01, ao primeiro período de coleta; e 18, ao ano de 2018. A segunda sentença foi salva, então, com a seguinte denominação: S2-1TRC-01-18; a terceira, S3-1TRC-01-18, e assim por diante. Esse procedimento foi realizado para todas as sentenças do CE.
- f. Na subpasta das sentenças limpas, adotamos o mesmo procedimento do item e, apenas acrescentando a letra L à denominação de cada sentença, para identificar que os arquivos já estão prontos para o processamento pela ferramenta.
- g. Em seguida, reunimos as sentenças limpas de cada Turma Recursal em dois arquivos, um para cada ano. Para a 1ª Turma Recursal Cível, por exemplo, cada arquivo recebeu a seguinte denominação: 1TRC-2018-REUNIDAS e 1TRC-2019-REUNIDAS. O mesmo procedimento foi adotado para as demais Turmas Recursais Cíveis.

Agora, passaremos a expor os procedimentos de coleta dos *corpora* de contraste CCA e CCB.

5.2.2 *Corpus* de Contraste CCA

O *corpus* de contraste CCA, como já mencionamos, é constituído de textos de dois jornais tradicionais: Zero Hora e Folha de São Paulo⁵⁴.

a) **Folha de São Paulo**

⁵⁴ Somos assinantes dos dois jornais e, por isso, conseguimos acesso às matérias deles extraídas.

Os textos da Folha de São Paulo são referentes às edições *on-line*⁵⁵ desse jornal do período de fevereiro de 2019 a junho de 2020. A coleta se deu em dias variados e os textos foram extraídos do menu “Notícias mais lidas”⁵⁶ de cada edição consultada.

Os textos da Folha de São Paulo foram armazenados numa pasta denominada FSP (Folha de São Paulo), com subpastas relativas ao período de extração. Por exemplo, a subpasta FSP-02-20 indica que os arquivos são do mês de fevereiro de 2020; a subpasta FSP-03-20 indica que os arquivos são mês de março de 2020.

Para os arquivos limpos (sem formatação), criamos uma pasta denominada FSP-L (Folha de São Paulo Limpos), para a qual adotamos o mesmo procedimento em relação à pasta FSP e às suas subpastas.

Cada arquivo foi salvo com a denominação como a que segue: FSP-20-02-19. FSP indica Folha de São Paulo, e 20-02-19 indica a data da edição. Assim, cada arquivo tem a sua data respectiva após FSP.

Para os arquivos limpos (sem formatação), adotamos o mesmo procedimento, apenas acrescentando a letra L após cada denominação.

O total de palavras dos arquivos da Folha de São Paulo, pela contagem do editor de texto Word, ficou em 549.991 palavras. Todos os arquivos, juntos e submetidos à ferramenta AntConc 3.2.1w., geraram 544.246 *tokens* e 30.333 *types*.

b) Zero Hora

Os textos do Jornal Zero Hora, tal como os textos da Folha de São Paulo, foram extraídos de modo *on-line* da versão acessível do site GZH⁵⁷. As matérias extraídas foram relativas a várias seções deste jornal, como Informe Especial, Política, Economia, Campo e Lavoura, Esporte, etc.

Os textos do Jornal Zero Hora foram armazenados numa pasta denominada ZH (Zero Hora), com subpastas relativas ao período de extração. Por exemplo, a subpasta ZH-02-20 indica que os arquivos são mês de fevereiro de 2020; a subpasta ZH-03-20 indica que os arquivos são mês de março de 2020.

Cada arquivo foi salvo com a denominação como a que segue: ZH-20-02-19. ZH indica Zero Hora, e 20-02-19 indica a data da edição. Assim, cada arquivo tem a sua data respectiva após ZH. Para os arquivos limpos (sem formatação), criamos uma pasta

⁵⁵ Disponível em <https://www.folha.uol.com.br/>

⁵⁶ Escolhemos esta seção, porque, por serem “mais lidas”, são acessadas por um maior número de leitores.

⁵⁷ Disponível em: <https://flipzh.clicrbs.com.br/jornal-digital/pub/gruporbs/acessivel/>.

denominada ZH-L (Zero Hora Limpos), para a qual adotamos o mesmo procedimento em relação à pasta ZH e às suas subpastas.

O total de palavras dos arquivos do Jornal Zero Hora, pela contagem do editor de texto Word, ficou em 516.560. Todos os arquivos, juntos e submetidos à ferramenta AntConc 3.2.1w, geraram 507.075 *tokens* e 28.630 *types*.

5.2.3 *Corpus de Contraste CCB*

O *corpus* contraste CCB, como já mencionamos, é constituído de textos de dois jornais populares: Diário Gaúcho e Jornal Massa.

a) **Diário Gaúcho**

Os textos do Diário Gaúcho foram extraídos do *Site PorPopular*⁵⁸. Este *site*, coordenado pela Prof^a. Dra. Maria José Borcony Finnatto, tem por foco a

descrição e o estudo de padrões do vocabulário exibido por textos de jornais populares voltados para públicos de menor poder aquisitivo [...] caracterização do léxico e da feição da linguagem como um todo em um texto que é feito, em tese, de um modo mais simplificado, para ser compreendido com facilidade por pessoas de um determinado grupo social e econômico, com uma bagagem cultural mais ou menos tipificada e com um grau de escolaridade relativamente baixo.⁵⁹

Para o Diário Gaúcho, utilizamos textos relativos aos anos de 2008 e 2010⁶⁰, que, reunidos, geraram um arquivo com 502.093 palavras segundo o editor de texto Word. O mesmo arquivo submetido ao AntConc 3.2.1w gerou 496.712 *tokens* e 30.196 *types*.

Os textos do Diário Gaúcho foram armazenados numa pasta denominada DG (Diário Gaúcho), em 8 arquivos: DG_C1; DG_C2; DG_C3; DG_C4; DG_C5a; DG_C5b, DG_C6; e DG_C7.

⁵⁸ Disponível em: <http://www.ufrgs.br/textecc/porlexbras/porpopular/download_do_corpus.php>. Acesso em 20 fev.2018.

⁵⁹ Disponível em: <<http://www.ufrgs.br/textecc/porlexbras/porpopular/objetivos.php>>. Acesso em 10 ago. 2018.

⁶⁰ Ainda que possam ser textos mais antigos que os dos demais jornais, optamos por mantê-los para estudo, visto que representam uma época em que o editorial desse jornal era exclusivamente voltado para a redação de um jornal popular. Hoje, a equipe responsável pela redação do Diário Gaúcho é a mesma que redige os textos do jornal Zero Hora.

b) **Jornal Massa**

Os textos do Jornal Massa foram extraídos do *site*⁶¹ desse jornal. As matérias foram relativas a várias seções deste jornal de edições de agosto de 2019 a dezembro de 2020.

Os arquivos dessas edições foram armazenados numa pasta denominada JM (Jornal Massa), com subpastas relativas ao período de extração. Por exemplo, a subpasta JM-08-19 indica que os arquivos são do mês de agosto de 2019, a subpasta JM-09-19 indica que os arquivos são do mês de setembro de 2019, e assim sucessivamente.

O total de palavras dos arquivos do Jornal Massa, pela contagem do editor de texto Word, ficou em 503.018. Todos os arquivos, juntos e submetidos à ferramenta AntConc 3.2.1w, geraram 491.667 *tokens* e 27.990 *types*.

A seguir, no Quadro 5.6, mostramos a classificação dos nossos *corpora* de contraste conforme os critérios utilizados por Berber Sardinha (2004) exemplificados no Quadro 4.1 deste trabalho.

Quadro 5.6 – Caracterização dos Corpora de Contraste

MODO	Escrito	
TEMPO	Sincrônico e Contemporâneo	
SELEÇÃO	De amostragem/Estático	Amostra finita.
	Equilibrado	Os componentes são distribuídos em quantidades semelhantes.
CONTEÚDO	Língua Geral	Os textos representam uma ou mais variedades sociolinguísticas. O CCA – jornais de referência – apresentam um português mais culto. O CCB – jornais populares – apresentam um português mais popular.
	Regional ou dialetal	
AUTORIA	De língua nativa	Os autores são falantes nativos.
	Institucional	O texto é produzido por empresas jornalísticas.
FINALIDADE	De contraste	Os <i>corpora</i> são usados para fins de contraste com o <i>corpus</i> de estudo.

Fonte: Produzido pela autora com base em BERBER SARDINHA (2004, p. 20-2); DUARTE (2008, p.161-3).

⁶¹ Disponível em: < <http://edicaodigital.atarde.com.br/massa/>>. Acesso em: 12 ago. 2018.

5.2.4 *Corpus* de Referência 1 – CR1

O *Corpus* de Referência 1 foi por nós compilado com obras literárias brasileiras que estão em domínio público⁶². Acessamos o *site* Domínio Público e selecionamos algumas obras de autores dos séculos XIX e XX, tal como relacionamos no Quadro 5.2, no item 5.1.5.

Para salvarmos esses arquivos, criamos 1 pasta denominada CR1 – OBRAS LITERÁRIAS e, dentro dela, 11 subpastas assim denominadas: CR1-AAZEVEDO, para Aluísio Azevedo; CR1-ECUNHA, para Euclides da Cunha; e assim fizemos para todos os autores selecionados. Em cada uma dessas subpastas, colocamos os arquivos em .pdf referentes às obras do autor respectivo. Cada arquivo foi salvo com o nome da obra baixada.

Após, transformamos cada arquivo em .pdf baixado em formato .docx e, posteriormente, em formato .txt. Em seguida, em cada subpasta, reunimos os arquivos .txt de cada autor e salvamos em um arquivo só, como no exemplo referente ao autor Aluísio Azevedo: AZEVEDO-T, que corresponde a todas as obras desse autor num arquivo só.

Depois disso, criamos um arquivo OBRAS LITERÁRIAS-T, onde colocamos os arquivos agrupados de cada autor. Submetemos esse arquivo OBRAS LITERÁRIAS-T no AntConc 3.2.1w., que gerou o resultado de 2.460.268 *tokens* e 82.136 *types*.

Selecionamos 11 autores diferentes e um conjunto de textos de cada um deles que não fosse muito discrepante em termos de *tokens*. A ideia foi construir um *corpus* variado e representativo da literatura brasileira dos séculos XIX e XX. Na Tabela 5.5, abaixo, está a lista de todos os autores e obras, com os resultados individuais quanto a *tokens* e *types*, após submissão das obras reunidas de cada um deles ao AntConc 3.2.1w.

Tabela 5.5 – Número de *Tokens* e *Types* por Autor do CR1

Autor	<i>Tokens</i>	<i>Types</i>
Aluísio Azevedo	259.870	22.986
Euclides da Cunha	115.570	20.288
Franklin Távora	180.498	18.057
Inglês de Souza	107.103	14.407
Joaquim Macedo	277.446	21.309
José de Alencar	246.464	20.626

62 **Domínio público.** Disponível em: www.dominiopublico.gov.br.

Lima Barreto	294.286	26.856
Machado de Assis	480.197	25.216
Paulo Setúbal	257.301	25.551
Raul Pompeia	115.252	18.837
Visconde de Taunay	125.681	19.000
Total	2.459.668	233.133

Como se vê na Tabela 5.5, o autor que apresenta maior número de *tokens* é Machado de Assis – 480.197 -, e o que apresenta menor número é Inglês de Souza – 107.103. Tendo em conta a iconicidade de Machado de Assis na literatura brasileira e maior facilidade de acesso às suas obras, não consideramos inadequado deixar essa diferença. Até porque, o número de *types* desse autor é muito semelhante, por exemplo, ao de Paulo Setúbal, cujas obras reunidas geraram 257.301 *tokens*. Considerando a genialidade de Machado de Assis, tal dado nos permite afirmar que pouca variação vocabular não significa pobreza de estilo e de talento. Nesse sentido, em estudo em que analisam redações de vestibulares, Finatto, Cremonese e Azeredo (2008, p. 99) verificaram “que os textos mais bem sucedidos tendem a lidar com um universo menor de palavras diferentes e que a repetitividade não tende a ser, enfim, uma inimiga do bom texto”.

Em relação ao número de *types* entre os autores, a diferença parece ser menor: 14.407 para Inglês de Souza, que também apresentou menor número de *tokens*; e 26.856 para Lima Barreto.

Quanto ao número total de *types*, a diferença é bem acentuada – 233.133 para 82.136 -, mas é previsível. Os *types* de cada autor individualmente são as palavras diferentes que esse autor empregou em suas obras. Porém, essas palavras diferentes de cada autor podem também ter sido usadas pelos outros autores em suas respectivas obras. Desse modo, ao se juntarem os textos de todos os autores num arquivo só, essas palavras diferentes de cada autor deixam de ser exclusivas e, por isso, o número de *types* diminui no todo.

Tendo exposto os dados e a forma de compilação dos *corpora* utilizados em nossa pesquisa, a seguir, na Tabela 5.6, colocamos informações relacionadas ao número de *tokens* e *types* de cada um deles, de acordo com a ferramenta *AntConc3.2.1w* (ANTHONY, 2019).

Tabela 5.6 – Número de *Tokens* e *Types* do CE, dos *Corpora* de Contraste e de Referência

<i>CORPORA</i>		Quantidade de <i>Tokens</i>	<i>Types</i>
CE		542.836	14.665
CCA	FSP	544.246	30.333
	ZH	507.075	28.630
CCB	DG	496.712	30.196
	JM	491.667	27.990
CCP		684.799	32.138
CR1		2.460.268	82.136
CR2		3.896.392.719 ⁶³	

Na próxima seção trataremos das ferramentas a serem utilizadas para a análise dos *corpora*.

5.3 FERRAMENTAS UTILIZADAS NAS ANÁLISES

Em nosso estudo, para que pudéssemos estimar em que medida os textos das Sentenças judiciais são mais ou menos potencialmente adequados ao seu destinatário, valemo-nos das ferramentas criadas no âmbito da LCOMP ou Processamento da Linguagem Natural (PLN).

O surgimento da LCOMP está relacionado ao desenvolvimento, nas décadas de 1950 e 1960, de programas de tradução automática, e sua história está intrinsecamente ligada a desenvolvimentos na área da Inteligência Artificial. Mas o maior crescimento se deu a partir do final dos anos de 1970. As pesquisas na Linguística Computacional dependem de diferentes áreas da Linguística e da Informática, e somente a interação entre os estudiosos dessas duas ciências é que pode gerar bons resultados nesse campo do saber.

Finatto, Lopes e Ciulla (2015) confirmam o resultado positivo da parceria entre os cientistas da computação e os linguistas, afirmando que a LC é uma das áreas que se vale de ferramentas e métodos computacionais para descrever, de modo estatístico, extensas amostras de usos da linguagem natural. No que se refere ao trabalho dos linguistas, além de se valerem das informações extraídas pelas ferramentas do PLN,

⁶³ Registramos apenas o número de palavras, conforme observação da nota de rodapé 36.

podem contribuir “para a melhoria desses sistemas, com a consideração de regras linguísticas que possam ser a ele integradas” (FINATTO; LOPES; CIULLA, 2015, p. 47).

Quanto à ST, os resultados das pesquisas no âmbito da LCOMP, já desde os anos 60, têm aparecido sob

a forma de sistemas que geram versões mais simplificadas de textos, incluindo a produção de ferramentas capazes de indicar diferentes tipos de medidas de CT. Essas ferramentas também conseguem produzir diferentes tipos de representações esquemáticas do conteúdo de um texto ou de todo *corpus*. Esses sistemas de LCOMP, de base fundamentalmente estatística, conseguem inclusive reconhecer tipologias textuais e graus de complexidade a eles associadas”. (FINATTO, 2011, p. 34)

Muitos procedimentos de ST se encontram em meio a projetos relacionados com a LCOMP. É o caso, por exemplo, de dicionários e de bases de dados que sustentam uma série de ferramentas ou de *software* cujo objetivo é auxiliar, com sugestões, um redator a simplificar um texto automaticamente a partir de um texto-fonte. Segundo Finatto e Motta, (2019, p. 330),

resultados concretos desse tipo de pesquisa podem ser conferidos, por exemplo, no caso da língua inglesa, observando-se o sistema da *Wikipedia Simple*, disponível em https://simple.wikipedia.org/wiki/Main_Page. Para quem deseje conhecer uma ferramenta simplificadora on-line, sugerimos conhecer o *SIMPLISH*, um sistema gerador automático de textos simplificados em inglês disponível em: <https://www.simplish.org>.

No Brasil, há o Projeto PorSimples, que disponibilizava duas ferramentas: o SIMPLIFICA e o FACILITA. O primeiro, SIMPLIFICA, visava a auxiliar autores a adequarem seus textos ao público-alvo pretendido. Esse editor exibia possíveis pontos de complexidade léxica e sintática e sugeria alterações. O segundo, FACILITA, resumia e simplificava textos da Web automaticamente. Este Projeto PorSimples encontra-se relatado em Aluísio e Gasperin (2010). As ferramentas que integravam a base do funcionamento do SIMPLIFICA - que operam, no âmbito lexical, com dicionários simples=>complexo e, no âmbito sintático, com um analisador de padrão de frases - podem ser visualizadas em <http://www.nilc.icmc.usp.br/nilc/index.php/tools-and-resources?layout=edit&id=27>. Este sistema, atualmente, não está em funcionamento.

A Figura 5.16, a seguir, ilustra a interface do sistema SIMPLIFICA.

Figura 5.16 – Sistema SIMPLIFICA, do Projeto PorSimples do NILC-USP



Fonte: <http://www.nilc.icmc.usp.br/simplifica/>. Acesso em 27 jul. 2018.

Ainda conforme Finatto e Motta (2019, p. 331), pelo menos pela via das pesquisas em PLN e em Informática, várias tentativas em prol da AT têm sido construídas e colocadas em teste, as quais, em geral, poucos linguistas brasileiros conhecem ou citam em seus estudos. A despeito das qualidades, defeitos, críticas por causa de superficialidade e de quaisquer limitações derivadas de um “automatismo computacional” que possam ser evocadas após experiências com o manejo de sistemas e *softwares* antes citados, podemos nos questionar sobre qual teria sido a participação de linguistas e de terminólogos em meio a essas iniciativas e sobre como elas poderiam ser aperfeiçoadas. O fato é que alguém, principalmente pesquisadores da área de Informática, tem tentado fazer algo concreto, oferecendo-nos inclusive alguns recursos de acesso público e gratuito.

Para a concretização deste estudo, nos valem do sistema NILC-Metrix⁶⁴ (Coh-Metrix 3.0, sua nova versão de 2020). O NILC-Metrix agrupa as métricas desenvolvidas em mais de uma década no NILC, iniciadas com o Coh-Metrix-Port e tem um uso mais facilitado: “Refatorado por Sidney Leal (com base nos códigos de Carolina Scarton, Andre Cunha e Nathan Hartmann), apoiado por Magali Duran. Todos sob a orientação da professora Sandra Aluísio”.

O Coh-Metrix-Port, que deu origem ao NILC-Metrix, é uma adaptação do sistema Coh-Metrix, que significa *cohesion metrics*. Trata-se de uma ferramenta para análise de textos em inglês, disponível gratuitamente *on-line*:

⁶⁴ Disponível em: <http://fw.nilc.icmc.usp.br:23380/nilcmetrix>

Elaborada por pesquisadores da Universidade de Memphis, nos Estados Unidos (GRAESSER; McNAMARA; LOUWERSE; CAI, 2004), tem como propósito calcular índices de coesão e de coerência textual num amplo espectro de medidas lexicais, sintáticas, semânticas e referenciais com o fim de indicar a adequação de um texto a seu público-alvo (a “demanda cognitiva” e a legibilidade do texto). Também tem a função de apontar dados para identificar problemas textuais de ordem estrutural. (FINATTO, 2011, p. 39)

No Brasil, o Núcleo Interinstitucional de Linguística Computacional da USP – NILC – com o objetivo de construir um sistema para possibilitar que analfabetos funcionais, pessoas com problemas cognitivos como afasia ou dislexia, por exemplo, e crianças e adultos em fase de aprendizado da leitura tivessem acesso à informação, adaptou o Coh-Metrix em inglês para o português, o Coh-Metrix-Port. Esta adaptação ofereceu inicialmente, das 60 métricas do sistema em inglês, apenas 35. O acesso é gratuito e se dá no âmbito do Projeto PorSimples (Simplificação Textual do Português para Inclusão e Acessibilidade Digital).

Na ferramenta SIMPLIFICA, “o Coh-Metrix-Port é responsável por analisar a inteligibilidade do texto e classificá-lo conforme o nível de alfabetização necessário para compreendê-lo: rudimentar, básico ou pleno, segundo os critérios do INAF” (CUNHA, 2015, p. 54).

O Coh-Metrix-Port, em suas diferentes versões, tem sido usado como ferramenta de análise em muitos trabalhos relativos à CT. É o caso, por exemplo, das pesquisas por nós já mencionadas de Pasqualini (2012) e de Scarton e Aluísio (2010). Estas últimas avaliaram a primeira versão do Coh-Metrix-Port, comparando textos escritos para adultos, hipoteticamente mais complexos, e textos escritos para crianças, hipoteticamente menos complexos. Nessa pesquisa, as autoras utilizaram textos regulares do jornal Zero Hora – tidos como complexos – e da seção *Para seu filho ler* do mesmo jornal – tidos como simples -, e textos das revistas Ciência Hoje – tidos como complexos – e Ciência Hoje das Crianças – tidos como simples. A conclusão a que chegaram foi a de que, como as métricas do Coh-Metrix-Port discriminaram com precisão textos dedicados a crianças e a adultos, elas também poderiam apresentar esse desempenho em textos de outros domínios.

Outro trabalho a merecer destaque é o de Finatto *et al.* (2011), que utilizou as métricas do Coh-Metrix-Port para avaliar contrastivamente a inteligibilidade de textos de dois jornais de Porto Alegre: o Zero Hora, destinado a pessoas com mais escolaridade, e o Diário Gaúcho, destinado a pessoas menos escolarizadas. Como um dos resultados da pesquisa, tem-se a evidência de que a ferramenta Coh-Metrix-Port “tem a capacidade de

distinguir tipos de jornais, separando com uma boa *F-measure*, o jornal popular do tradicional” (FINATTO *et al.*, 2011, p. 8).

Nas conclusões deste artigo, as autoras sugerem a criação de uma ferramenta “de suporte à escrita, por exemplo, para jornalistas novatos elaborarem e identificarem textos mais ou menos adequados para uma ou outra classe de leitores” (*Ibidem*). Essa sugestão vem ao encontro de uma das possibilidades que nossa pesquisa pode gerar: os dados dela advindos poderão abastecer futuras ferramentas *on-line* com o objetivo de “traduzir” a linguagem jurídica para um português simplificado.

A seguir, então, descrevemos como se dá o acesso ao NILC-Metrix e os recursos que essa ferramenta disponibiliza.

Para acessar a página, deve-se digitar o seguinte endereço <http://fw.nilc.icmc.usp.br:23380/nilcmetrix>, onde surge a tela exposta na Figura 5.17:

Figura 5.17 – Tela Inicial do NILC-Metrix



Para submeter um texto novo, basta colar, no espaço com contorno vermelho na Figura 5.18 abaixo.

Figura 5.18 – Espaço para colocar o texto a ser analisado pelo NILC-Metrix



Após a ferramenta processar o texto, surge a tela da Figura 5.19. Nessa tela constam os resultados da análise. Esses resultados aparecem em uma tabela com 200 linhas. Cada linha apresenta 1 das 200 métricas utilizadas por essa ferramenta.

Figura 5.19 – Tela com os resultados da análise pelo NILC-Metrix

Grupo	Métrica	Descrição	Valor
1	Coesão Referencial adj_arg_ovf	Quantidade média de referentes que se repetem nos pares de sentenças adjacentes do texto	1.66667
2	Coesão Referencial adj_ov_ovf	Quantidade média de palavras de conteúdo que se repetem nos pares de sentenças adjacentes do texto	2.0
3	Coesão Referencial adj_stem_ovf	Quantidade média de radicais de palavras de conteúdo que se repetem nos pares de sentenças adjacentes do texto	2.33333
4	Coesão Referencial arg_ovf	Quantidade média de referentes que se repetem nos pares de sentenças do texto	1.66667
5	Coesão Referencial stem_ovf	Quantidade média de radicais de palavras de conteúdo que se repetem nos pares de sentenças do texto	1.5

Essas métricas são agrupadas pelo tipo de análise que fazem: 9 de coesão referencial, 11 de coesão semântica, 27 de complexidade sintática, 12 de conectivos, 4 de densidade de padrões sintáticos, 15 de diversidade lexical, 10 de frequência de palavras, 42 de informações morfossintáticas, 11 de informações semânticas, 12 de léxico temporal, 10 de medidas descritivas, 24 de medidas psicolinguísticas, 9 de simplicidade textual, 5 de índices de leituraabilidade.

Em cada linha, é colocado o número da linha, o grupo em que se insere a métrica, o nome da métrica, uma descrição explicando o que a métrica analisa e, por fim, na última coluna, é colocado o valor da métrica no texto processado.

Essas medidas possibilitam uma infinidade de análises e têm sido usadas em muitas pesquisas relacionadas à inteligibilidade textual. Porém, cumpre aqui citar o que diz Pasqualini:

As métricas calculadas pelas ferramentas Coh-Metrix e Coh-Metrix-Port, por si sós, não indicam o nível de complexidade de um texto. É na inter-relação entre as métricas que se encontra o melhor caminho na avaliação da complexidade. (PASQUALINI, 2012, p. 77)

Além disso, segundo Finatto e Evers (2016, p. 274), esses índices devem ser entendidos como uma referência, e não como um fim em si mesmos. E é esta postura que adotamos em nossas análises nesta pesquisa. Ou seja, com base nos índices relativos às métricas avaliadas pela ferramenta NILC-Matrix nos diferentes textos de nossos *corpora*, procuraremos descrever eventuais semelhanças ou discrepâncias entre eles, de modo a identificar traços colaboradores para uma maior CT.

Além do NILC-Matrix, utilizamos outros recursos informatizados, como o AntConc 3.2.1w⁶⁵, *software* livre que fornece listas de palavras, concordâncias e outras utilidades, e o Sketch Engine⁶⁶, que, conforme já mencionamos, fornece lista de palavras por critérios variados – frequência, classe gramatical etc. -, *thesaurus* de distribuição, concordâncias, extração de terminologias, criação e comparação de *corpora*, entre outros.

Fromm, Grama, Beilke e Santos (2020) fazem um estudo comparativo entre as funcionalidades do *Sketch Engine* e do *Wordsmith Tools*, outra ferramenta utilizada para processamento e análise de textos. Concluíram que ambos os programas são eficazes naquilo a que se propõem e que a escolha dependerá dos propósitos da pesquisa e do próprio pesquisador.

Em nossa pesquisa optamos por utilizar o AntConc 3.2.1w para a contagem de *tokens* e *types*, pois o *Sketch Engine* não faz essa diferenciação. Entretanto, este último oferece mais recursos que o AntConc 3.2.1 w, por ser uma plataforma paga. Além disso, o CR2 que utilizamos nas comparações de nossos *corpora* é disponibilizado nessa plataforma. Por isso, também utilizamos o *Sketch Engine* em nossas análises.

Na próxima seção apresentamos estudos parciais realizadas durante a pesquisa.

⁶⁵ **AntConc 3.2.1w**. Disponível em <http://www.laurenceanthony.net/software.html>

⁶⁶ **Sketch Engine**. Disponível em <https://the.sketchengine.co.uk>

6 ESTUDOS PARCIAIS 1

Tendo apresentado o tema de nossa tese, sua justificativa, os fundamentos teóricos que a embasam, os materiais e os métodos de que utilizamos para executá-la, passamos, nesta seção, a apresentar os estudos-pilotos que realizamos durante a pesquisa. Esses estudos correspondem a ensaios de aproximação ao nosso objeto, o todo do texto das Sentenças dos JECs.

Iniciamos nossa apresentação com o Estudo-Piloto 1, que foi nossa primeira experiência concreta com o tema de nossa pesquisa. Foi, porém, como veremos, bem importante para verificar o que poderíamos aplicar, ou não, na totalidade das análises dos nossos *corpora*.

6.1 ESTUDO-PILOTO 1 – PRIMEIRA EXPERIÊNCIA COM AS MÉTRICAS DA FERRAMENTA COH-METRIX-PORT

Neste primeiro Estudo-Piloto (MOTTA, 2018), fazemos uma análise contrastiva entre uma amostra de textos das Sentenças dos JECs com amostras dos textos dos jornais de nossos *corpora* de contraste, quais sejam: os jornais tradicionais Zero Hora, Folha de São Paulo (CCA); e os jornais populares Diário Gaúcho e Jornal Massa (CCB). Para tanto, utilizamos algumas das métricas selecionadas para as análises desenvolvidas na dissertação de mestrado de Mesquita (2011), pesquisa que tratou da inteligibilidade textual da Coleção Cadernos Didáticos de EJA, destinada a jovens e adultos egressos do ensino regular.

A seguir, no item 6.1.1, especificaremos a constituição das amostras dos *corpora* que dele fazem parte e, no item 6.1.2, comentaremos os dados das análises realizadas pela ferramenta *Coh-Matrix-Dementia*⁶⁷ relativamente às métricas selecionadas.

6.1.1 Constituição dos *corpora* do Estudo-Piloto 1

⁶⁷ Como o Coh-Matrix-Port não estava disponível, utilizamos o Coh-Matrix-Dementia (<http://143.107.183.175:22380/>). Essa ferramenta foi criada por Cunha (2015) que acrescentou 25 métricas ao Coh-Matrix-Port para ajudar na detecção precoce de problemas relacionados ao declínio nas habilidades de linguagem.

Para a constituição dos *corpora* deste Estudo-Piloto, adotamos os seguintes procedimentos:

- a. Selecionamos quatro Sentenças assim denominadas: S1TRC011-16, S2TRC011-16, S3TRC011-16 e S4TRC011-16 (S = sentença; 1TRC = Primeira Turma Recursal Cível, 2TRC = Segunda Turma Recursal Cível, 3TRC = Terceira Turma Recursal Cível, 4TRC = Quarta Turma Recursal Cível; 01 = primeiro período; 1 = primeira sentença; 16 = ano de 2016⁶⁸).
- b. Colocamos estas Sentenças num arquivo único denominado SP1, na pasta ESTUDO-PILOTO.
- c. Realizamos a limpeza no arquivo SP1, ou seja, retiramos os nomes das partes, dos julgadores e qualquer outra referência que pudesse trazer alguma identificação pessoal. Retiramos também outros sinais de formatação que pudessem impedir o processamento pela máquina.
- d. Salvamos, com a extensão .txt, o arquivo limpo, com 8.294 palavras, segundo o AntConc 3.2.1w., como SP1-L.
- e. Selecionamos um conjunto de textos do Jornal Zero Hora que gerou um arquivo, segundo o AntConc 3.2.1w., de 8.616 palavras.
- f. Denominamos o arquivo da Zero Hora como ZH-PILOTO e o salvamos, na pasta ESTUDO-PILOTO, subpasta ZH-PILOTO, com a extensão .txt.
- g. Selecionamos um conjunto de textos do Jornal Folha de São Paulo que gerou um arquivo de 8.125 palavras, segundo o AntConc 3.2.1w.,.
- h. Denominamos o arquivo da Folha de São Paulo como FSH-PILOTO e o salvamos, na pasta ESTUDO-PILOTO, subpasta FOLHA-PILOTO, com a extensão .txt.
- i. Estes *corpora* foram reunidos num conjunto chamado neste Estudo-Piloto de Caso A.
- j. Selecionamos, dos *corpora* do CCB, um conjunto de textos do Jornal Diário Gaúcho que gerou um arquivo de 8.335 palavras, segundo o AntConc 3.2.1w..
- k. Denominamos o arquivo do Diário Gaúcho como DG-PILOTO e o salvamos, na pasta ESTUDO-PILOTO, subpasta DG-PILOTO, com a extensão .txt.
- l. Selecionamos um conjunto de textos do Jornal Massa que gerou um arquivo de 8.313 palavras, segundo o AntConc 3.2.1w..
- m. Denominamos o arquivo do Massa como Massa-PILOTO e o salvamos, na pasta ESTUDO-PILOTO, subpasta MASSA-PILOTO, com a extensão .txt.
- n. Estes *corpora* foram reunidos num conjunto chamado neste Estudo-Piloto de Caso B.

Observamos que, neste estudo, os textos referentes à FSP foram extraídos do *site* da Linguateca-CETENFolha⁶⁹, onde, após preenchermos um formulário de identificação,

⁶⁸ Iniciamos selecionando sentenças de 2016 e 2017 para o nosso CE, porém, posteriormente, por problemas de acesso a uma quantidade maior de textos, optamos por selecionar sentenças de 2018 e 2019.

⁶⁹ **Linguateca-CETENFolha**. Disponível em: <http://www.linguateca.pt/CETENFolha/>

recebemos uma identificação e uma senha para acessarmos a página de *downloads*⁷⁰. Esses textos fazem parte da Coleção CHAVE, criada pelo Linguateca em cooperação com a organização do CLEF - *Cross- Language Education and Function*⁷¹. Em relação à Folha de São Paulo, integram essa Coleção CHAVE as edições completas dos anos de 1994 e 1995. Para o nosso estudo, foram utilizados textos do ano de 1994.

Os textos da ZH foram extraídos de um bloco de 80 textos dos anos de 2006 e 2007, com um total de 55.528 palavras, utilizado pelas pesquisadoras Maria José B. Finatto, Carolina E. Scarton, Amanda Rocha e Sandra Aluísio em artigo em que procuraram contrastar “a inteligibilidade de jornais populares e de jornais tradicionais brasileiros a partir de atributos coesivos, vocabulares e sintáticos, incluindo elipses” (FINATTO *et al.*, 2011, p. 49). Este conjunto de textos pertence ao Projeto PorSimples, do NILC da USP de São Carlos e nos foi cedido para a pesquisa⁷².

Os textos do DG, extraídos do *site* PorPopular, são referentes a meses variados do ano de 2013, e os relativos ao Jornal Massa são referentes a maio de 2011.

A seguir, colocamos a Tabela 6.1 com a quantidade de palavras (*tokens*), segundo o AntConc 3.2.1w., dos *corpora* deste Estudo-Piloto.

Tabela 6.1 – Quantidade de Palavras dos Corpora no Estudo-Piloto 1

	CORPUS DE ESTUDO	CASO A		CASO B	
		ZH	FSP	DG	MASSA
Quantidade de palavras	8.294	8.616	8.125	8.335	8.313

6.1.2 Processamento e Análise dos Dados

Conforme já mencionado, para este primeiro Estudo-Piloto, utilizamos algumas das métricas do Coh-Metrix-Dementia selecionadas para as análises desenvolvidas na dissertação de mestrado de Mesquita (2011). Essa foi uma pesquisa que tratou da inteligibilidade textual da Coleção Cadernos Didáticos do EJA, destinada a jovens e adultos egressos do ensino regular. Após a submissão de seus textos ao sistema Coh-Metrix-Port, esta autora comparou os resultados obtidos com os do Índice Flesch referentes aos mesmos textos.

⁷⁰ Disponível em: <http://www.linguateca.pt/chave/download/>

⁷¹ Disponível em: www.clef-campaign.org

⁷² Gentilmente cedidos pela Profa. Dra. Maria José Bocorny Finatto.

Vale ressaltar, conforme mencionamos na seção 5.3, que hoje dispomos de uma versão atualizada desse sistema, com muito mais facilidade para uso *on-line*, o Coh-Metrix 3.0, disponível em: <http://fw.nilc.icmc.usp.br:23380/cohmetrixport>. Apesar da nova versão (de 2020), o relato do nosso estudo piloto envolve, assim, uma versão anterior (de 2015). Entretanto, vários elementos com que já lidamos permanecem inalterados, como o Índice Flesch e outras contagens.

O Índice Flesch, conforme já mencionamos, é tido por muitos como uma fórmula superficial, pois considera apenas o número de palavras por sentença e o número de letras ou sílabas por palavra. Apesar disso, considerado válido como estimativa, foi incorporado ao Coh-Metrix-Port e também ao Coh-Metrix-Dementia e seguiu no Coh-Metrix 3.0, pois, grosso modo, ele ajuda a indicar a potencial complexidade na tarefa da leitura. Além disso, segundo Scarton e Aluísio (2010, p. 15), é uma métrica de inteligibilidade adaptada para o português que incorpora o conceito de séries escolares brasileiras.

Diante dessas especificidades do Índice Flesch, Mesquita (2011) o comparou com o resultado das outras métricas analisadas em seu *corpus*, para verificar em que medida esse índice refletiria, tomado isoladamente, a complexidade dos textos.

As métricas selecionadas por esta autora foram: *Índice Flesch; Sentenças por parágrafo; Sílabas por palavra de conteúdo; Número de palavras por sentenças; Incidência de negações; Frequência de Palavras de Conteúdo; Número de verbos por número de sentenças (cláusula por sentença); Incidência de conectivos; Palavras antes dos verbos principais; Referência anafórica adjacente; e Sobreposição de argumentos adjacentes*. Tais métricas, com os seus respectivos índices aplicados ao nosso CE, são apresentadas na Tabela 6.2, a seguir. Esses índices são o resultado do processamento de nosso CE pela ferramenta *Coh-Metrix-Dementia*.

Tabela 6.2 – Índices das Métricas do Coh-Metrix *Dementia* para o Estudo-Piloto 1

MÉTRICAS	CORPUS DE ESTUDO		CASO A		CASO B	
	SP1	SP2*	ZH-P	FSP-P	DG-P	MASSA-P
Índice Flesch	34,591	32,666	42,173	47,566	52,027	50,551
Cláusula por sentença	2,237	2,502	1,894	1,951	1,800	2,022
Sentenças por parágrafo	2,032	1,692	1,660	2,066	2,189	2,263
Palavras por sentença	22,059	25,288	17,442	16,298	14,101	17,033

Sílabas por palavra de conteúdo	2,943	2,955	2,872	2,829	2,785	2,733
Palavras antes dos verbos principais	4,005	4,767	4,755	4,575	4,618	5,144
Incidência de Negações	10,731	11,244	7,984	11,216	12,019	11,535
Incidência de Conectivos	65,831	67,593	73,121	80,994	85,135	81,844
Sobreposição de arg. adjacentes	1,056	1,186	0,395	0,398	0,169	0,546
Referência anafórica adjacente	0,184	0,231	0,124	0,240	0,151	0,238
Frequência de palavras de conteúdo	629308,386	449198,991	519824,356	436470,307	502725,080	523343,165

Fonte: Produção da autora com base nos índices calculados pelo Coh-Matrix-Dementia

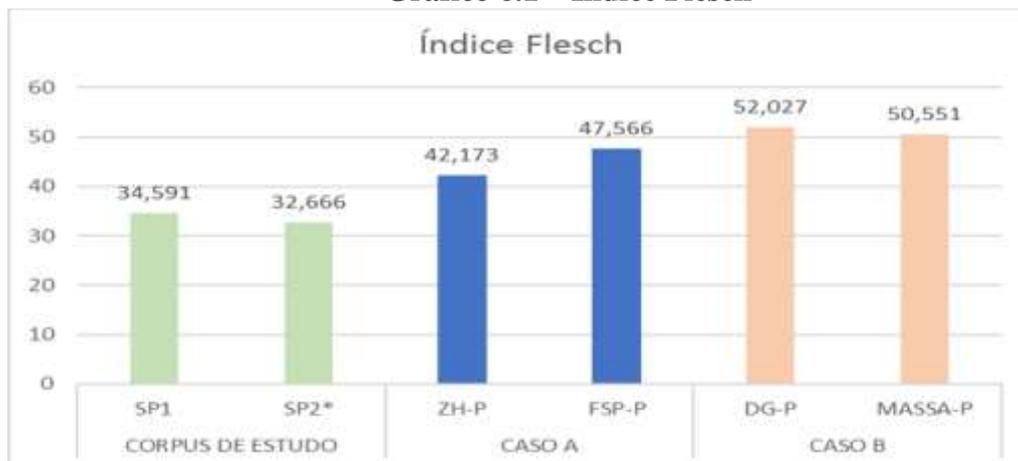
Na Tabela 6.2, acima, há uma coluna com SP2 (Sentenças-Piloto 2), não prevista na constituição dos nossos *corpora* do Estudo-Piloto. Como as Sentenças do nosso *corpus* apresentam muitas ementas, conforme explicamos na descrição do CE, e essas ementas, em seus cabeçalhos, apresentam muitas frases nominais - às vezes uma palavra só -, imaginamos que isso poderia alterar os índices desses textos. Poderia alterar principalmente os índices relativos ao número de palavras por sentença, que, inclusive, é considerado para o cálculo do Índice Flesch, para a aferição da inteligibilidade textual. Diante disso, como essas ementas são citações dentro do texto, e não o texto em si, fizemos um teste, criando este outro arquivo – SP2 –, sem o cabeçalho das ementas, para verificar se as diferenças entre o SP1 seriam significativas. Tais situações procuraremos descrever a seguir, quando colocarmos os resultados de nossas análises.

A seguir, então, em forma de itens, explicamos o que medem cada uma dessas métricas selecionadas por Mesquita (2011) e fazemos uma aplicação delas aos nossos *corpora*:

1) Índice Flesch

Abaixo, no Gráfico 6.1⁷³, apresentamos os números relativos ao Índice Flesch de nossos *corpora*.

⁷³ Os gráficos apresentados neste estudo-piloto foram gerados no Programa Excel com os dados advindos da ferramenta Coh-Matrix-Dementia.

Gráfico 6.1 – Índice Flesch

Como já mencionamos, o Índice Flesch considera o número de palavras por sentença e o número de letras ou sílabas por palavra. Quanto maior o número neste índice, mais fácil é considerado o texto. No Gráfico 6.1, o número acima de cada coluna corresponde ao Índice Flesch de cada *corpus*. E os números do eixo vertical do gráfico dão o indicativo da ordenação por dificuldade de cada *corpus* em relação ao Índice Flesch. No caso, o número menor indica o mais complexo, e o número mais alto o mais fácil.

Os demais gráficos apresentados neste Estudo-Piloto adotarão esses critérios para as métricas a que se referem.

Diante disso, os textos campeões em termos de facilidade neste quesito são os do Caso B, textos dos jornais populares Diário Gaúcho e Massa. Os textos do CE apresentaram-se mais difíceis, ficando os do SP2, sem os cabeçalhos das ementas, com o índice menor que o SP1. Isso vem a confirmar nossa suspeita de que a presença do cabeçalho das ementas poderia alterar a legibilidade desses textos. Tal mudança surpreendeu, pois, havendo mais frases nominais, a tendência é diminuir a extensão das frases e, conseqüentemente, aumentar o Índice Flesch. E não foi isso o que ocorreu. É preciso ponderar, porém, que a extensão das frases é apenas uma das variáveis consideradas pelo Índice Flesch, como já mencionamos.

2) Número de verbos por sentenças/Cláusula por Sentença

Mesquita (2011) explica que esse índice foi calculado com base em outros dois índices – *número de sentenças* e *incidência de verbos*. Ela dividiu o número de verbos pelo número de sentenças para identificar se o que a ferramenta define como sentença – que é um período, pois está entre pontos -, corresponderia a um período simples ou

composto. Ela explica que, se o índice fosse 2,2, por exemplo, poderia significar, em princípio, a presença de períodos compostos de duas orações, correspondendo cada uma delas a um verbo, visto que a ferramenta não distingue verbos de locuções verbais. Ela justifica este cálculo porque, em tese, períodos compostos são mais complexos do que períodos simples.

No que se refere ao nosso estudo, como nossa ferramenta não foi o Coh-Metrix-Port, mas o Coh-Metrix-Dementia, verificamos que uma das métricas inseridas por Cunha foi a **Cláusulas por Sentença**, que calcula o número médio de cláusulas por sentença, em que **cláusula** é caracterizada pela presença de um sintagma verbal. Para ilustrar o cálculo, o Manual do Coh-Metrix-Dementia traz o seguinte exemplo: "A mulher que eu vi usava um chapéu vermelho. O seu chapéu era muito bonito. A primeira sentença possui 2 cláusulas e a segunda sentença possui uma cláusula. Assim, o valor da métrica é 1,5" (CUNHA, 2015).

Pelo cálculo dessa métrica, quanto maior for o índice, maior é a incidência de verbos por sentença. Havendo maior incidência de verbos por sentença, pode-se inferir que haja um maior número de orações em cada sentença. Havendo maior número de orações, o período se torna mais complexo.

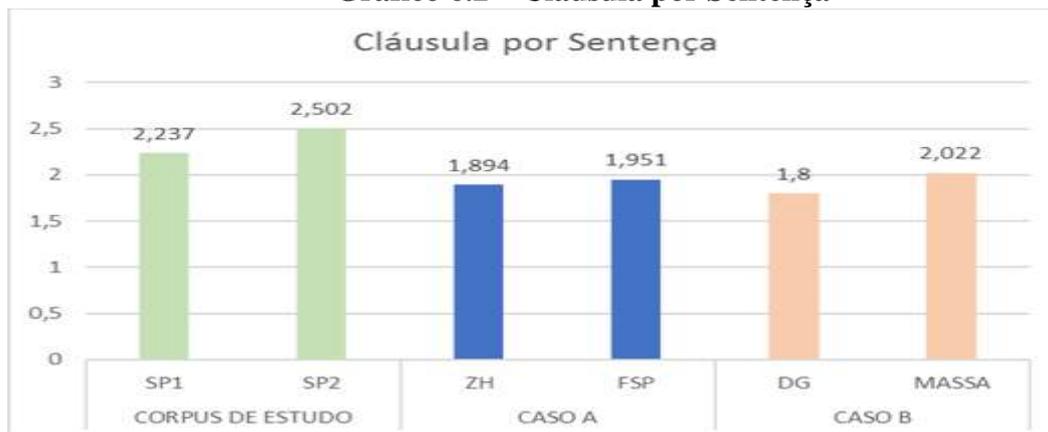
Cabem aqui duas observações. A primeira é o fato de a ferramenta não distinguir verbos de locuções verbais, o que influi no cálculo das orações. Para testar essa situação, colocamos na ferramenta o seguinte trecho, onde constam duas locuções verbais, com dois verbos cada.

O Brasil demorou a entender que precisava rever suas políticas.

A incidência de verbos ficou em 400, que reflete o cálculo $4/0,010$ em que 0,010 corresponde ao total de palavras da frase - 10 - dividido por 1000, e 4 ao número de verbos do texto. Ou seja, a ferramenta apenas considera o número de verbos isoladamente, sem distinguir eventuais locuções verbais presentes nos textos.

A segunda é quanto à não possibilidade de distinguir se o período é composto por coordenação ou subordinação, pois, em princípio, períodos compostos por subordinação são mais complexos que os compostos por coordenação.

De todo modo, cremos que esta métrica **Cláusula por Sentença** já faz o cálculo realizado por Mesquita (2011) nas suas análises. E, no que diz respeito aos dados dos nossos *corpora*, temos o que se pode ver no Gráfico 6.2, a seguir:

Gráfico 6.2 – Cláusula por Sentença

Pelo que vemos no Gráfico 6.2, os textos que se mostraram mais complexos quanto a esse aspecto foram os do CE, seguidos dos do Jornal Massa. E novamente os do SP2 se mostraram mais difíceis.

3) Número de sentenças por parágrafos

Mesquita (2011) afirma que o número de sentenças por parágrafo pode indicar se os parágrafos são longos ou curtos, o que pode estar relacionado ao grau de complexidade. Porém, cabe observar a orientação encontrada em muitos manuais de redação quanto a se preferir frases mais curtas a longas. Inclusive no *Federal Plain Language Guidelines* de 2011, temos o seguinte trecho: "Sentenças mais curtas também são melhores para transmitir informações complexas. Elas dividem as informações em unidades menores e mais fáceis de processar"⁷⁴ (PLAIN LANGUAGE ACTION AND INFORMATION NETWORK, 2011, p. 56, tradução nossa).

Assim, se um parágrafo é redigido com mais frases curtas, ele terá mais sentenças por parágrafo, mas nem por isso será mais complexo. Cremos que, neste aspecto, o que mais importa é a relação entre o **número de palavras por sentença** e o **número de sentenças no parágrafo**.

Abaixo, na Tabela 6.3, seguem os índices de nossos *corpora* em relação a esses aspectos.

⁷⁴ "Shorter sentences are also better for conveying complex information; they break the information up into smaller, easier-to-process units."

Tabela 6.3 – Índices Sentenças por Parágrafo e Palavras por Sentença

MÉTRICAS	CORPUS DE ESTUDO		CASO A		CASO B	
	SP1	SP2	ZH	FSP	DG	MASSA
Sentenças por parágrafo	2,032	1,692	1,660	2,066	2,189	2,263
Palavras por sentença	22,059	25,288	17,442	16,298	14,101	17,033

Fazendo-se o cálculo de multiplicação do **número de sentenças por parágrafo** pelo **número de palavras por sentença** para cada *corpus*, temos os seguintes dados, conforme Gráfico 6.3, a seguir:

Gráfico 6.3 – Sentenças por Parágrafo x Palavras por Sentença

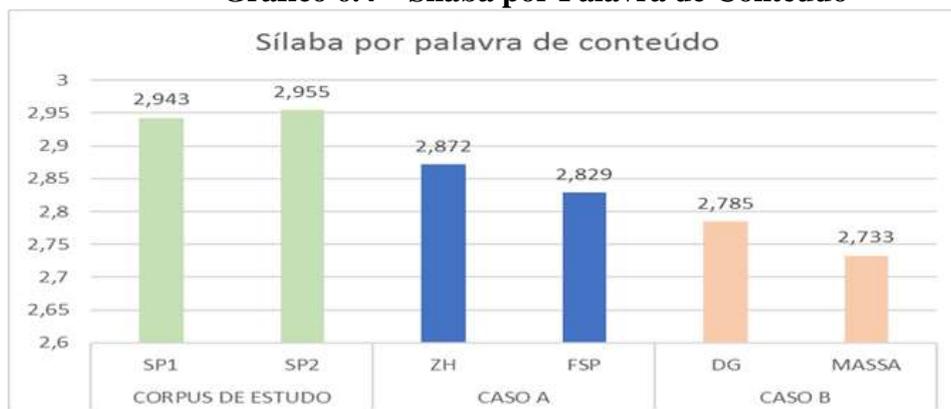
É possível verificar que os textos que apresentam os parágrafos com maior número de palavras e, em princípio, mais complexos, seriam os do CE, seguidos dos textos do Jornal Massa.

4) Sílabas por palavra de conteúdo

Mesquita (2011) justifica a adoção dessa métrica para suas análises porque analfabetos funcionais apresentam dificuldade na decodificação do sinal gráfico. Diz ela que leitores iniciantes precisam do apoio da linearidade das cadeias das letras para chegar ao sentido das palavras. Diz, ainda, que as palavras de conteúdo⁷⁵ desempenham um papel fundamental na compreensão leitora e que pesquisas apontam que os leitores fixam 80% do movimento dos olhos nesse tipo de palavra, o que não ocorre com as palavras funcionais.

Quanto a esse aspecto, os dados dos nossos *corpora* são os que constam no Gráfico 6.4, a seguir:

⁷⁵ Palavras de conteúdo ou palavras lexicais são palavras carregadas de significado, em princípio, independente, como os substantivos, adjetivos, advérbios e verbos.

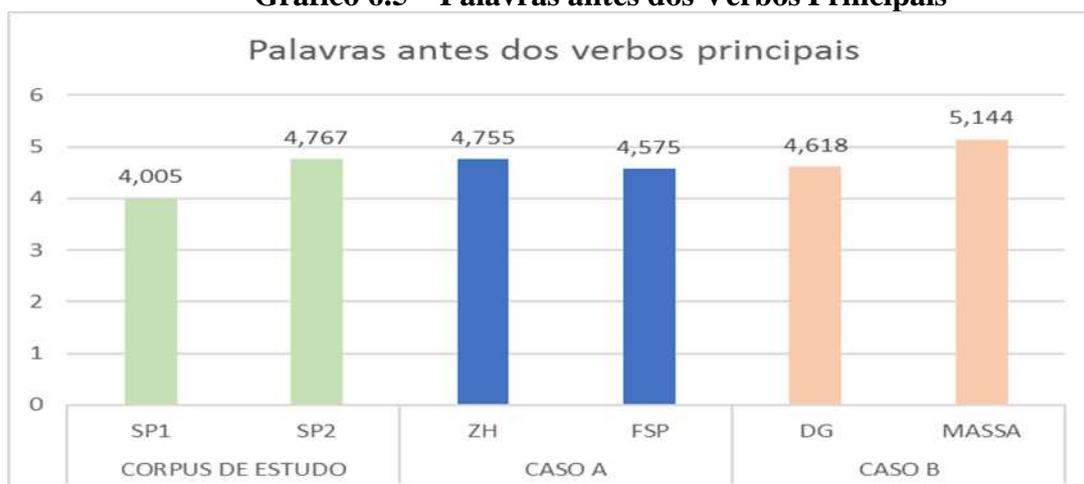
Gráfico 6.4 – Sílabas por Palavra de Conteúdo

Novamente, quanto a essa métrica, os textos que se apresentaram mais difíceis foram os do CE. Os do Caso B, nesta métrica, seguiram a lógica esperada, ou seja, de serem os mais acessíveis.

5) Palavras antes dos verbos principais

Em relação a essa métrica, Mesquita (2011) justifica a sua seleção porque, segundo a documentação do Coh-Metrix, é um bom índice para avaliar a carga de memória de trabalho relacionada à manutenção de informação relativa ao sujeito, quando do processamento pelo verbo. Aluísio e Scarton (2010, p. 53) afirmam que sentenças com muitas palavras antes do verbo principal são mais complexas, pois sobrecarregam a memória de trabalho dos leitores. Nesse artigo, elas comentam que os textos de jornais reescritos para crianças apresentaram uma diminuição nessa métrica, o que elas consideraram um bom resultado, pois seria um facilitador da leitura.

Vejamos, agora, no Gráfico 6.5, como ficaram os nossos *corpora* quanto a esse aspecto.

Gráfico 6.5 – Palavras antes dos Verbos Principais

Como vemos no Gráfico 6.5, os textos que apresentaram maior índice neste aspecto foram os do Jornal Massa, seguidos dos textos da Zero Hora, ficando em terceiro lugar os textos das SP2. Além disso, o Diário Gaúcho vem em quarto lugar, ficando à frente da Folha de São Paulo.

Tais dados surpreendem, porque se espera que os jornais populares (Caso B), tendo em vista seu público-alvo, sejam de uma leitura mais acessível, o que não parece ocorrer neste caso, de acordo com o critério mencionado.

Cabe observar também a diferença nos índices entre as Sentenças SP1 e as Sentenças SP2. As Sentenças SP1 ficaram, neste quesito, mais acessíveis. Cremos que a explicação para isso está no fato de que este arquivo de Sentenças contém os cabeçalhos das ementas. Tais cabeçalhos, como já mencionamos, contém muitas frases nominais e palavras isoladas por ponto, o que é considerado uma sentença. Desse modo, havendo mais sentenças sem verbos, esse índice tende a diminuir.

6) Incidência de negações

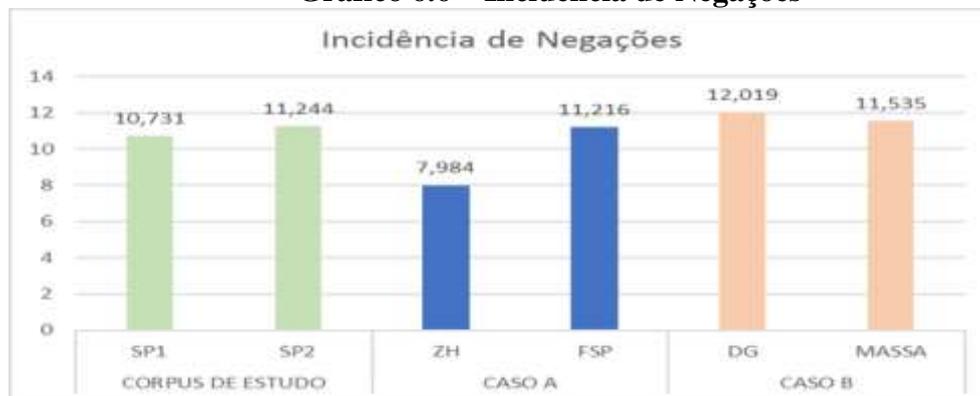
No que se refere às negações, Mesquita (2011) afirma que a incidência maior de negações parece dificultar a compreensão:

Resultados de experimentos envolvendo identificação de figuras a partir de sentenças afirmativas e negativas indicaram que estas são mais difíceis em termos de processamento do que aquelas (Carpenter & Just, 1975; Clark & Chase, 1972; Trabbasso *et al.*, 1971 apud Tettamanti *et al.*, 2008). Conforme os autores esclarecem neste artigo, ‘a compreensão de sentenças negativas demanda maior tempo de processamento, e estas apresentam maior percentagem de erros, conforme resultados de testes que envolvem pareamento de frases e figuras’. Além disso, estudos sobre acessibilidade de informação mencionada no escopo da negação, envolvendo tarefas de reconhecimento de palavras ou tarefas de *priming* mostraram que os tempos gastos para responder

eram significativamente mais longos para os itens com negações (MacDonald and Just, 1989; Kaup, 2001; Kaup and Zwaan, 2003 apud Tettamanti *et al.*, 2008). (MESQUITA, 2011, p. 55).

A seguir, no Gráfico 6.6, os dados de nosso *corpus* relativo a essa métrica.

Gráfico 6.6 – Incidência de Negações



Pelo Gráfico 6.6, os textos do Diário Gaúcho apresentaram um maior índice de negações, seguido do Jornal Massa. Ou seja, os jornais populares apresentaram-se mais complexos neste quesito. Tendo em vista que Pasqualini (2012) deixou de usar esta métrica por haver problemas na medição do Coh-Metrix-Port, averiguamos, em cada *corpus* de nosso estudo, a presença dos itens relativos à negação e verificamos que os índices de nossos *corpora* estavam corretos.

7) Incidência de Conectivos

Mesquita (2011) afirma que, em princípio, um texto com presença de conectivos facilita o processamento, porque as relações semânticas entre as sentenças são indicadas por esses elementos. Ela cita Perfetti (2005), para quem os leitores pouco proficientes fixam-se mais nas palavras isoladamente do que na busca pela coerência do texto e que leitores habilidosos são melhores nas atividades de inferência de conexões do texto. Diante disso, afirma que a presença de conectivos, auxiliando no estabelecimento das relações semânticas interfrásticas, é fundamental para leitores pouco proficientes. E essa presença explícita do conectivo, mesmo tornando a frase por vezes mais extensa com processos de coordenação ou subordinação, é um facilitador da leitura, porque as relações semânticas do trecho ficam mais evidentes.

Quanto a esta métrica, os dados dos nossos *corpora* são os que constam no Gráfico 6.7, a seguir:

Gráfico 6.7 – Incidência de Conectivos

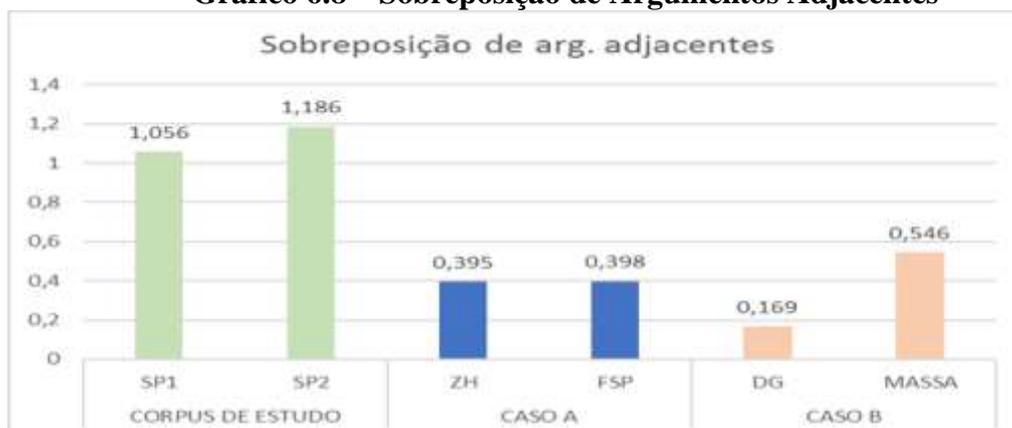
Os textos que se mostraram mais difíceis relativamente a esta métrica foram os do CE, pois o índice é menor. E os mais fáceis, com índices mais elevados, são os do Caso B.

8) Sobreposição de argumentos adjacentes

Em relação a esta métrica, Mesquita explica que

números mais elevados podem ser tomados como facilitadores da leitura, pois, em princípio, quando há sobreposição de argumentos, a informação é retomada, facilitando a manutenção da referência e o processo de progressão temática. (MESQUITA, 2011, p. 57)

Os dados dos nossos *corpora* em relação a esta métrica podemos ver no Gráfico 6.8, a seguir:

Gráfico 6.8 – Sobreposição de Argumentos Adjacentes

Pelo Gráfico 6.8 acima, os textos com maiores índices são os do CE, que se apresentam, neste quesito, menos difíceis. E a diferença com relação aos outros *corpora*

é significativa, considerando-se, por exemplo, o índice dos textos do Diário Gaúcho, 0,169.

Uma possível explicação para isso pode estar no fato de que o CE é composto por textos especializados do mundo jurídico. Nesse âmbito, há orientações para que, em prol da precisão terminológica, caso haja necessidade, se repitam as palavras. É o que se vê no seguinte trecho de Ruy Rosado de Aguiar Júnior, Desembargador do PJRS e Ministro do Superior Tribunal de Justiça aposentado:

Se for preciso repetir as palavras, que estas sejam repetidas, isto é, não é preciso invocar sinônimos para expressar as ideias que estão na ementa. Especialmente, não se deve inventar: denúncia é denúncia, e não peça-ovo nem peça-pórtico; petição inicial é petição inicial, e não portal. Se tivermos de dizer dez vezes petição inicial, diremos sempre petição inicial, pois assim diz a lei, e é assim que devemos escrever. (AGUIAR JÚNIOR, 2008).

Pode-se aplicar aqui também a análise feita por Pasqualini (2011), em artigo de que trata da CT em dois textos técnico-científicos da área de Pediatria e suas respectivas traduções para o inglês. Nesse trabalho, Pasqualini, após relacionar as métricas do Coh-Matrix-Port referentes a incidência de sintagmas nominais, incidência de anáforas e a proporção *types/tokens*, assim se manifesta:

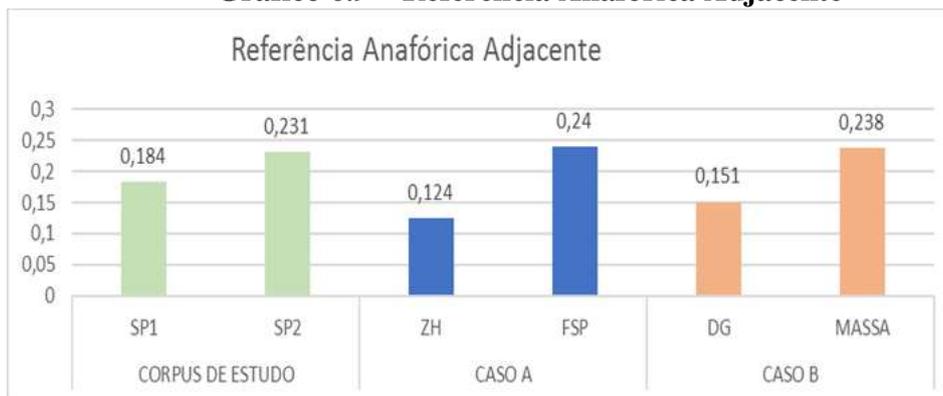
uma variação pequena de vocabulário (um índice *types/token* baixo), uma grande incidência de sintagmas e uma baixa presença de anáforas indica um elevado grau de repetição vocabular, o que, para um texto técnico-científico, talvez seja um indício de legibilidade e de densidade terminológica. (PASQUALINI, 2011, p. 15)

9) Referência Anafórica Adjacente

Relativamente a esta métrica, Mesquita (2011) explica, com base no exemplo do Manual do Coh-Matrix-Port, que, quanto mais alto for o seu valor, mais difícil será, em tese, o texto, porque haverá, em princípio, mais candidatos a serem retomados pela anáfora.

Foi verificado que, quando há dois possíveis referentes para um pronome, esses leitores levam mais tempo para atribuir o pronome ao referente pretendido, ao menos quando o referente não corresponde ao primeiro nome da frase precedente. (Mesquita, 2011, p. 58)

A seguir, no Gráfico 6.9, estão os dados dos nossos *corpora* em relação a esta métrica:

Gráfico 6.9 – Referência Anafórica Adjacente

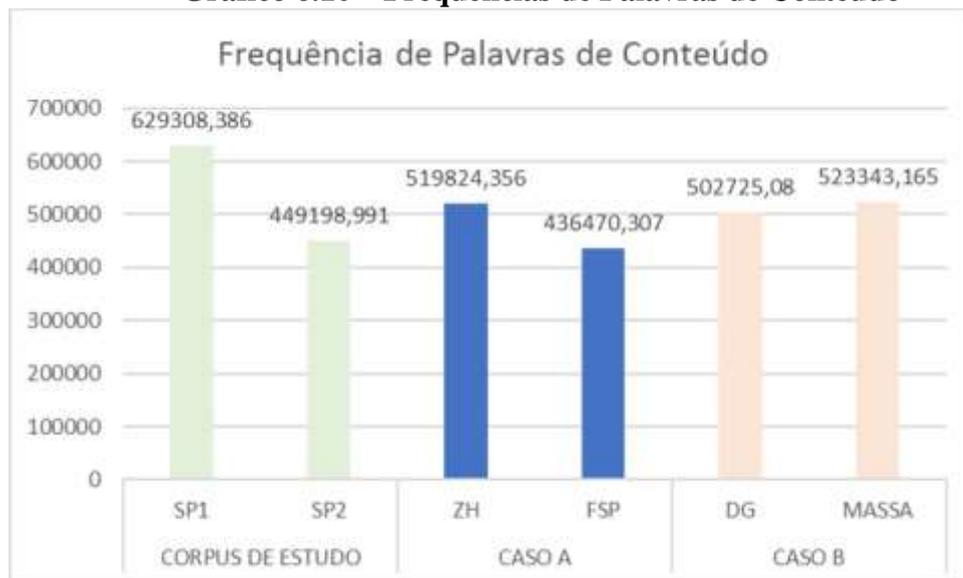
Podemos ver, no Gráfico 6.9, que os textos com índice maior relativamente a esta métrica foram os dos Jornais Folha de São Paulo e Massa, revelando-se, portanto, mais difíceis que os demais quando a este aspecto.

10) Frequências de Palavras de Conteúdo

Mesquita (2011) explica que esta métrica indica a média de todas as frequências de palavras de conteúdo e que a presença de palavras pouco frequentes em um texto pode também contribuir para sua complexidade.

Na pesquisa de Aluísio e Scarton (2010), essas autoras observaram que os textos reescritos, mais fáceis, apresentaram um índice maior de frequência de palavras de conteúdo que os textos originais. Com isso, concluíram que esses textos reescritos apresentaram mais palavras frequentes do que os textos originais.

No Gráfico 6.10, a seguir, estão os dados dos nossos *corpora* em relação a essas métricas:

Gráfico 6.10 – Frequências de Palavras de Conteúdo

Como podemos ver no Gráfico 6.10, os textos que se apresentaram menos complexos neste quesito foram os da SP1. Aqui também parece calhar a observação feita quanto à repetição de palavras na métrica relativa à sobreposição de argumentos adjacentes.

A seguir, com base nas análises a que procedemos acima, apresentamos a Tabela 6.4, onde, excluindo o Índice Flesch e considerando as demais métricas analisadas, apresentamos a hierarquia de legibilidade de cada um de nossos *corpora*, tal como fez Mesquita (2011) em relação aos 13 cadernos da Coleção Cadernos Didáticos do EJA.

Tabela 6.4 – Hierarquia de Legibilidade dos Corpora

MÉTRICAS	CORPUS DE ESTUDO		CASO A		CASO B	
	SP1	SP2*	ZH-P	FSP-P	DG-P	MASSA
Cláusula por sentença	2	1	4	5	6	3
Sentenças por parágrafo x Palavras por sentença	1	2	6	4	5	3
Sílabas por palavra de conteúdo	2	1	3	4	5	6
Palavras antes do verbo principal	6	2	3	5	4	1
Incidência de negação	5	3	6	4	1	2
Incidência de conectivos	1	2	3	4	6	5
Sobreposição de arg. adjacentes	5	6	2	3	1	4
Anáfora adjacente	4	3	6	1	5	2
Frequência de palavras de conteúdo	1	5	3	6	4	2
MÉDIA	3	2,777 (mais difícil)	4	4	4,111 (mais fácil)	3,111

Para compreensão da Tabela 6.4, acima, a numeração de 1 a 6 corresponde a uma ordem de complexidade, sendo 1 o número atribuído ao texto mais complexo e 6 ao mais simples. Os números foram atribuídos conforme os índices colocados na Tabela 6.2. Após cada arquivo receber um número de acordo com o critério acima, fizemos a média desses números, colocada na última linha da tabela.

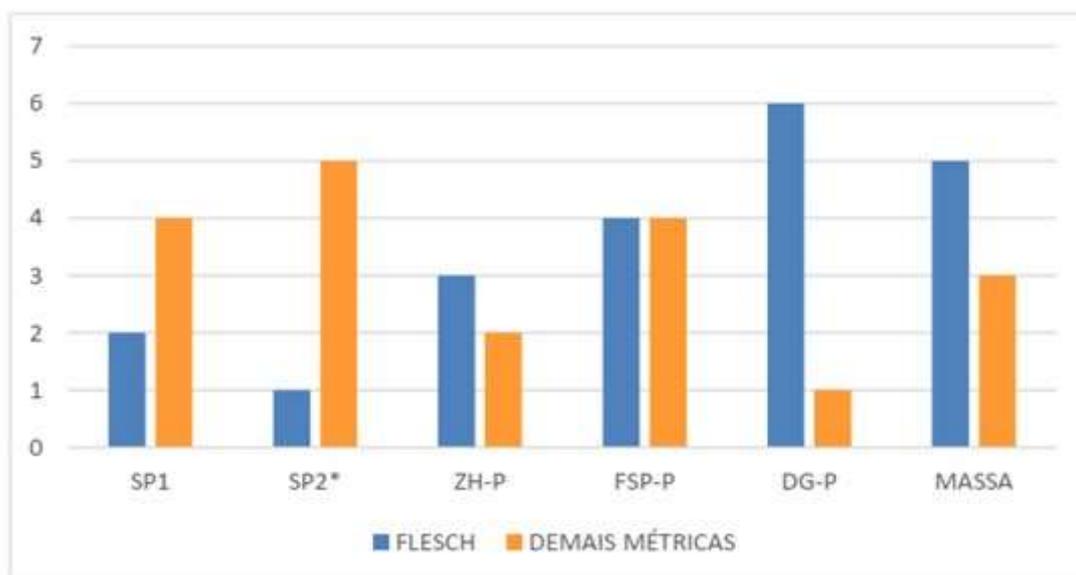
E, a seguir, colocamos a Tabela 6.5, em que comparamos essa ordenação com o Índice Flesch de cada um dos arquivos, para verificar se este índice reflete, isoladamente, a complexidade dos textos.

Tabela 6.5 – Comparação das Médias da Tabela 6.4 com o Índice Flesch

MÉTRICAS	CORPUS DE ESTUDO		CASO A		CASO B	
	SP1	SP2*	ZH-P	FSP-P	DG-P	MASSA
Flesch	34,591	32,666	42,173	47,566	52,027	50,550
ORDENAÇÃO DO FLESCH	2	1 (mais difícil)	3	4	6 (mais fácil)	5
MÉDIA DA TABELA 6.4	3	2,777	4	4	4,111	3,111
ORDENAÇÃO DA MÉDIA DAS DEMAIS MÉTRICAS	4	5 (mais fácil)	2	2	1 (mais difícil)	3

Com os dados relativos à ordenação do Índice Flesch e das demais métricas em conjunto, conforme a Tabela 6.5, foi possível se chegar ao Gráfico 6.11.

Gráfico 6.11 – Índice Flesch x Demais Métricas



Tendo em conta que o número menor do eixo vertical do Gráfico 6.11 indica maior complexidade, e o número maior, menor complexidade, os textos mais fáceis em relação ao Índice Flesch são os dos jornais populares do Caso B. Porém, há uma enorme discrepância em relação à média das outras métricas analisadas. Vemos que o DG, pela média das métricas analisadas, ficou como o texto mais difícil (coluna laranja no número 1).

Como já mencionamos, o Índice Flesch considera apenas o número de palavras por sentença e o número de letras ou sílabas por palavra, e por isso é tido como superficial. Pelo gráfico acima, é possível inferir que efetivamente assim seja.

Por sua vez, as métricas analisadas neste Estudo-Piloto são de ordem **lexical** – *sentença por parágrafo, sílabas por palavra de conteúdo, palavras por sentença, frequência de palavras de conteúdo* –, **sintática** – *incidência de negação, incidência de conectivos, número de palavras antes dos verbos* – e **semântica** – *anáforas, sobreposição de argumentos*. Ou seja, as métricas analisadas contemplam um espectro maior de atributos do texto que o Índice Flesch e podem, por isso, trazer uma imagem mais real acerca dos textos analisados. Porém, é preciso observar que as métricas *sentença por parágrafo versus palavras por sentença e sílabas por palavra de conteúdo* conferem com os resultados trazidos pelo Índice Flesch quanto a serem os textos do CE mais complexos que os demais.

Diante de todas as medições feitas neste Estudo-Piloto, saltam aos olhos os índices dos textos jornalísticos, principalmente os dos jornais populares. Considerando que a atribuição de complexidade feita pelas métricas analisadas em nosso Estudo-Piloto, em comparação com os dados do Índice Flesch, seja mais próxima da realidade dos textos de nossos *corpora*, **causa-nos surpresa o fato de um jornal popular – DG – ficar como o mais complexo entre os textos analisados**. Como já referimos, os jornais populares são destinados às classes C, D e E, com baixa escolaridade e pouca experiência em leitura. E, por terem este leitor como destinatário de seus textos, em princípio, apresentariam uma linguagem mais simples, o que parece não estar acontecendo.

Nesse sentido, julgamos pertinente trazer as observações de Finatto *et al.* (2011), em artigo de que trataram de características do jornalismo popular. O objetivo desse trabalho foi verificar quais características o texto do jornal popular – no caso, o Diário Gaúcho – se diferenciaria do texto do jornal tradicional – no caso, o Zero Hora.

Segundo esses pesquisadores, os jornais populares brasileiros tinham incrementado o índice de leitura de grandes parcelas da população urbana de baixa renda.

Para tanto, afirmam, com base em Amaral (2006, p. 39), que uma pesquisa do IBOPE revelou que o DG foi responsável pela elevação do índice de leitores da região metropolitana de Porto Alegre para o maior de todo o Brasil. Observam também que “[o] grau de instrução dos leitores do DG está assim distribuído: 60% com Ensino Fundamental; 34% com Ensino Médio e 6% com Ensino Superior” (FINATTO *et al.*, 2011, p. 50).

Nesse artigo, para analisar os textos dos jornais, serviram-se das métricas do Coh-Matrix-Port e acrescentaram mais cinco métricas relativas à presença de elipses. E, pelos resultados obtidos, verificaram que o

DG, popular, exhibe, comparado ao tradicional ZH, similar uso de elipses. Portanto, se elipses equivalassem a texto mais complexo e até sofisticado em termos de elaboração, poder-se-ia detectar justamente aí um dos traços de um gênero novo, popular e ao mesmo tempo complexo” (FINATTO *et al.*, 2011, p. 55-6).

Como vimos, os dados obtidos de nosso Estudo-Piloto parecem também favorecer a conclusão de estarmos diante de um “gênero novo, popular e ao mesmo tempo complexo”.

Antes de concluirmos esta exposição, cumpre-nos referir que este Estudo-Piloto foi nossa primeira experiência de análise textual com base em métricas de uma ferramenta como o Coh-Matrix. Os dados aqui brevemente apresentados nos instigaram a estender as comparações realizadas neste Estudo-Piloto para conjuntos de textos maiores, como apresentamos no Capítulo 8 deste trabalho. Cremos que, com isso, poderemos melhor compreender as razões pelas quais este “gênero novo” que representa os jornais populares é, em tese, mais bem compreendido do que as Sentenças judiciais.

6.2 ESTUDO-PILOTO 2 – ÍNDICES LEXICAIS E PROPOSTAS DE SIMPLIFICAÇÃO DE TRECHOS DE UMA AMOSTRA DE SENTENÇAS X UMA AMOSTRA DO DG X CORPOP

Neste Estudo-Piloto pretendemos verificar e estimar alguns índices lexicais de uma amostra de nosso *corpus* de Sentenças e propor, quando necessário, alternativas de reescritas mais simples, vistas como processos de simplificação textual e/ou tradução intralinguística. Como já assinalamos, por *inteligibilidade textual* entendemos, preliminarmente, um conjunto de características linguísticas estruturais do texto escrito,

potencialmente associadas a maiores facilidades ou dificuldades de compreensão do texto por parte de diferentes perfis de leitores.

Nossa descrição e análise tiveram como parâmetro um *corpus* de textos do Diário Gaúcho e dados do CorPop. O CorPop, vale lembrar, é um *corpus* de referência do PPE, desenvolvido no trabalho de Pasqualini (2018), disponível para consulta em: <http://www.ufrgs.br/textecc/porlexbras/corpop/index.php>.

6.2.1 Constituição dos *corpora* do Estudo-Piloto 2

Os *corpora* utilizados neste Estudo-Piloto são:

- a) as Sentenças dos JECs utilizadas no Estudo-Piloto 1;
- b) o conjunto de textos do DG utilizados também no Estudo-Piloto 1; e
- c) o CorPop.

6.2.2 Processamento e Análise dos Dados

Para a descrição e análise dos *corpora* em contraste, valemo-nos das ferramentas computacionais AntConc 3.2.1w⁷⁶, do Sketch Engine⁷⁷ e do Coh-Metrix-Port⁷⁸.

Neste estudo, primeiramente geramos no AntConc 3.2.1w a lista de palavras dos textos das Sentenças judiciais e dos materiais do jornal DG. Por *palavra*, nesse tipo de ferramenta computacional, entende-se a *palavra gráfica*, composta por uma sequência de caracteres separada por um espaço em branco. Posteriormente, submetemos todos os textos sob exame ao sistema Coh-Metrix-Port. Das 48 métricas disponibilizadas por essa ferramenta, selecionamos para este estudo apenas o Índice Flesch (doravante IF). Por óbvio, a adoção de somente um índice não esgota o quadro de fatores envolvidos na CT, mas nossa ideia foi a de uma aproximação inicial para ajuste de procedimentos em maior escala, no âmbito da tese como um todo.

Na Tabela 6.6, a seguir, seguem os dados de nossos *corpora* relativos a esses índices e também aos dados do AntConc 3.2.1w.

⁷⁶ AntConc 3.2.1w. Disponível em <http://www.laurenceanthony.net/software.html>

⁷⁷ Sketch Engine. Disponível em <https://the.sketchengine.co.uk>.

⁷⁸ Coh-Metrix-Port. Disponível em <http://143.107.183.175:22680/>.

Tabela 6.6 – Dados dos corpora extraídos do AntConc e do Coh-Metrix-Port

	SENTENÇAS	DG
<i>tokens</i> (total de palavras)	8.149	8.278
<i>types</i> (palavras diferentes)	2.228	2.293
Relação <i>types/tokens</i> - TTR (variedade do vocabulário)	0,27	0,28
Densidade Semântica	1,648	1,533
IF	34,53	53,93

A Tabela 6.6 mostra, além do IF, a relação *types/tokens*. Não utilizamos os dados do Coh-Metrix-Port nesse cálculo, porque ele considera, nesta avaliação, apenas as palavras de conteúdo. Os números relativos à TTR apresentados na Tabela 6.6 são resultado de um cálculo manual entre o número de palavras diferentes (*types*) e o número de total de palavras (*tokens*).

Os resultados mostram que os textos são muito semelhantes quanto à TTR. Essa relação considera a razão entre o total de palavras (*tokens*) e as palavras diferentes (*types*) do texto. Quanto maior o número dessa relação, mais variado é o léxico do texto. Quanto menor, mais repetitivo, o que parece ser a situação dos nossos corpora. No caso das Sentenças, com um índice de 0,27, uma possível explicação estaria no fato de serem textos especializados, nos quais, em prol da precisão terminológica, há muita repetição de palavras. No caso do DG, o índice de 0,28 pode indicar que uma menor variação no vocabulário facilitaria a compreensão do público-alvo deste jornal. De acordo com Finatto *et al.* (2011), os leitores do DG, em sua maioria – 60% – têm apenas o Ensino Fundamental.

Em relação à DS, quanto maior for o resultado numérico, mais denso semanticamente é o texto. Assim se dá porque essa métrica mede a relação entre palavras lexicais, como *processo* (um substantivo), e palavras funcionais, como *de* (preposição). Quanto mais palavras lexicais apresentar o texto, mais palavras o leitor deverá conhecer para compreender o que nele está escrito:

Textos que tenham índices altos de densidade semântica são aqueles que exigem maior conhecimento previamente adquirido dos leitores e que exijam maior esforço cognitivo para interpretar a significação não só das palavras, mas também da relação entre as palavras. (SILVA, 2018, p. 121)

Nesse quesito, as Sentenças, com 1,648, mostram-se mais complexas que os textos do DG, com 1,533. Quanto ao IF, os textos das Sentenças, com 34,53, são considerados difíceis, e os textos do DG, com 53,93, fáceis.

Após, procedemos à comparação, inicialmente, entre a lista de palavras das Sentenças com a lista de palavras do DG; em seguida, entre a lista de palavras das Sentenças e o CorPop; e, por fim, entre a lista de palavras do DG e o CorPop. Essa comparação foi realizada com a lista lematizada das palavras. O processo de lematização consiste em reduzir palavras variáveis à correspondente forma canônica, como, por exemplo, os verbos no infinitivo, os substantivos e adjetivos no singular, etc. Para a lematização das palavras de nossos *corpora*, utilizamos a ferramenta Sketch Engine e depois fizemos uma triagem com os mesmos critérios adotados para a lematização do CorPop⁷⁹.

Depois da lematização, a lista de palavras das Sentenças ficou com 1299 itens, e a do DG com 1321 itens. O CorPop, por sua vez, apresenta 5235 lemas.

A inclusão do CorPop nessas comparações se deve ao fato de essa lista ser um *corpus* de referência do PPE que pode servir como baliza para a estimativa da complexidade lexical de um texto. Assim, palavras dos nossos *corpora* que não se encontram na lista do CorPop podem ser consideradas “como potencialmente complexas ou passíveis de simplificação” (PASQUALINI, 2018, p. 92).

A seguir, na Tabela 6.7, seguem os dados relativos a essas comparações.

Tabela 6.7 – Comparação entre os Lemas dos *Corpora*

	Número Total de Lemas	Lemas inexistentes no CorPop		Lemas inexistentes no DG	
		Nº	%	Nº	%
Sentenças	1297	494	38,09	947	73,01
DG	1321	307	23,24		

Na Tabela 6.7, do total de lemas das Sentenças, 38,09% não aparecem na lista do CorPop e 73,01% não aparecem na lista do DG. Isso significaria, em tese, que cerca de 38% do vocabulário das Sentenças é potencialmente desconhecido para o nosso leitor-

79 Retirada de nomes próprios; de nomes de cidades, países e localidades; de preposições; de pronomes pessoais; de advérbios comuns, como “sim”, “não”, “mal”, “nem”; de artigos definidos e indefinidos; de pronomes demonstrativos; de interjeições; de meses do ano e dias da semana; de prefixos soltos (pré, pós, etc.); de siglas; de numerais cardinais e ordinais; de algumas palavras e/ou regionalismos.

alvo. Cabe salientar, porém, que a proporção maior de lemas das Sentenças inexistentes em relação ao DG pode estar relacionada ao fato de essa lista ser quase 4 vezes menor que a lista do CorPop (o DG tem 1321 lemas, e o CorPop, 5235). Essa comparação, assim, exigiria uma melhor ponderação matemática.

De qualquer sorte, levando em conta que os lemas ausentes do CorPop podem ser considerados potencialmente complexos, as Sentenças se apresentam mais difíceis que os textos do DG nesse quesito. No Quadro 6.1, a seguir, apresentamos alguns exemplos desses lemas das Sentenças e do DG ausentes no CorPop.

Quadro 6.1 – Exemplos de Lemas das Sentenças e do DG Inexistentes no CorPop

Lemas das Sentenças Inexistentes no CorPop	Lemas do DG Inexistentes no CorPop
<i>acostar, avença, coligir, consumerista, culpabilidade, dissuasório, egrégio, ensejar, excludente, expurgo, fulcro, hipossuficiente, inadimplemento, incolumidade, indébito, indubitável, instância, jurisprudência, jurisdição, perpetrar, probatório, prolação, sinistralidade, solver, sucumbência, tutela.</i>	<i>açougueiro, alagamento, ambulatorial, barral, bueiro, ecografia, encomenda, engarrafamento, fossa, fratura, hidráulica, informatizar, lajota, manicure, máximo, mureta, parafuso, patrola, pavimentação, proeza, sacola, saibro, semáforo, viário, vistoria, zelar.</i>

As palavras do DG exemplificadas no Quadro 6.1 referem-se ao cotidiano do cidadão, tais como *açougueiro, alagamento, engarrafamento, manicure, etc.* Essas palavras, portanto, apesar de não constarem no CorPop, fazem parte do dia a dia do cidadão brasileiro, independentemente do grau de instrução, e podem ser consideradas simples.

Por outro lado, palavras como *avença, fulcro, hipossuficiente*, que aparecem na lista de palavras das Sentenças exemplificadas no Quadro 6.1, ainda que possam ser empregadas em outros textos que não os do âmbito do Direito, elas ocorrem quase que exclusivamente em textos jurídicos. Algumas podem ser consideradas termos, como é o caso, por exemplo, de *consumerista, indébito e jurisprudência*. Outras, no entanto, muitas vezes representam preciosismos e jargões jurídicos que conformariam o tão falado “juridiquês”, ainda que isso possa ser entendido apenas como traços de erudição, como seria o caso das palavras *indubitável, perpetrar* ou *solver*. Essas podem ser, então, razões para que tais itens lexicais não constem no CorPop.

Diante disso, retomamos aqui que, se tais palavras não se encontram no CorPop, podem ser consideradas potencialmente complexas. É preciso, porém, ponderar aqui a questão temporal dos *corpora*, visto que, de acordo com o momento, muitas palavras

passam a ser mais usadas que em outros. É o que ocorreu no caso da pandemia do coronavírus. Palavras como *pandemia*, *quarentena*, e os estrangeirismos *home office*, *lockdown* passaram a ser mais frequentes na língua portuguesa. Mesmo não se encontrando no CorPop, não significa que elas sejam palavras mais complexas ou menos frequentes na língua como um todo. Apenas estão mais na moda do que em outras épocas.

De todo modo, tendo em conta que um dos objetivos deste estudo é propor, quando necessário, reescritas mais adequadas ao cidadão destinatário das Sentenças dos JECs – contexto em que tais palavras se encontram -, selecionamos algumas das palavras das Sentenças exemplificadas no Quadro 6.1, os trechos em que ocorrem e realizamos os procedimentos explicados na seção a seguir.

6.2.3 Propostas de Reescrita e/ou Tradução Intralinguística

Entre as palavras das Sentenças ausentes no CorPop, selecionamos para este experimento *incolumidade* e *sinistralidade*. Primeiramente, procuramos, na Sentença de onde foram extraídas, se havia alguma explicação, alguma paráfrase para cada uma delas. Nada encontramos nesse sentido. Após, selecionamos o parágrafo onde se encontravam e propusemos alternativas de reescritas simplificadas, vista aqui como um procedimento de tradução intralinguística.

Tais reescritas procuraram seguir as orientações de autores como DuBay (2004), que embasam as orientações constantes nos manuais (*Guidelines*) do *Plain Language*. Servimo-nos também dos resultados dos estudos de Silva (2018), mencionados no item 4.5, no qual trazemos dados de vários estudos relacionados à ATT.

Ao comparar o desempenho das medidas utilizadas em sua pesquisa⁸⁰, Silva (2018) concluiu que as que apresentaram melhores resultados em termos de redução da CT foram (em ordem): simplificação lexical, redução de adjetivos e redução de informação.

Outro trabalho que também nos serviu de orientação nesse procedimento de simplificação foi o de Carpio (2017), igualmente mencionado no item 4.5 de nosso trabalho. Nessa pesquisa com textos sobre doenças respiratórias associadas ao trabalho,

⁸⁰ a) Índice Flesch (IF); b) análise semântica latente (ASL); c) relação *type-token* (TT); d) densidade semântica (DS); e) incidência de substantivos (S); f) incidência de verbos (V); g) incidência de adjetivos (ADJ); h) incidência de advérbios (ADV); e i) incidência de pronomes (P).

com base nas características potencializadoras de CT observadas, a pesquisadora apresentou reescritas mais simples e, em tese, mais afeitas à compreensão do seu público-alvo. Para essas reescritas simplificadas, esta autora adotou estratégias como “o texto deve ser breve e todas as informações apresentadas devem ser concisas e literais [...] evitar o uso desnecessário de léxico especializado [...] e incluir uma explicação clara e sucinta em caso de que a terminologia seja inevitável” (CARPIO, 2017, p. 50).

A seguir, então, ilustramos esses procedimentos com as duas palavras selecionadas.

Incolumidade

Esta palavra, no âmbito do Direito, significa a condição de estar livre de dano. No caso do sintagma terminológico *incolumidade subjetiva*, como o que ocorre no trecho em questão, o significado é a condição de estar livre de dano em algum aspecto subjetivo da pessoa, como a sua imagem, a sua intimidade etc. Não há, na totalidade do texto, explicação sobre o que essa expressão significa. Assim, inicialmente, numa estratégia de simplificação lexical, apresentamos, ao trecho original em que ela está inserida, uma reescrita em que colocamos um sinônimo mais frequente na língua portuguesa: a palavra *intimidade*, que tem 30 ocorrências no CorPop. É o que podemos ver no Quadro 6.2, a seguir.

Quadro 6.2 – Reescrita 1 - Incolumidade

Trecho Original	Reescrita 1
A lide versa a respeito de defeito do serviço, de acordo com artigo 14, <i>caput</i> , do Código de Defesa do Consumidor, em que o objeto da relação de consumo apresenta falha que afeta a segurança legitimamente esperada quanto à incolumidade subjetiva , do consumidor ou do seu patrimônio.	A lide versa a respeito de defeito do serviço, de acordo com artigo 14, <i>caput</i> , do Código de Defesa do Consumidor, em que o objeto da relação de consumo apresenta falha que afeta a segurança legitimamente esperada quanto à intimidade do consumidor ou do seu patrimônio.
IF = 10.330	IF = 17.53
DS = 1.136	DS = 1.043
TTR = 0,766	TTR = 0,744

Abaixo de cada trecho, no Quadro 6.2, constam os índices relativos às métricas por nós consideradas neste Estudo-Piloto. O IF aumentou na Reescrita 1, o que significa

um texto menos complexo que o original, porém ainda difícil. Os outros índices diminuíram, o que também revela que a Reescrita 1 apresenta-se menos complexa que o trecho original.

Num segundo momento, propusemos a Reescrita 2 a partir da Reescrita 1. Servindo-nos da estratégia de simplificação por explicação, acrescentamos uma explicação para o que seria a intimidade do consumidor, como aparece grifado no Quadro 6.3.

Quadro 6.3 – Reescrita 2 - Incolumidade

Reescrita 1	Reescrita 2
A lide versa a respeito de defeito do serviço, de acordo com o artigo 14, caput, do Código de Defesa do Consumidor, em que o objeto da relação de consumo apresenta falha que afeta a segurança legitimamente esperada quanto à intimidade do consumidor ou do seu patrimônio.	A lide versa a respeito de defeito do serviço, de acordo com o artigo 14, caput, do Código de Defesa do Consumidor, em que o objeto da relação de consumo apresenta falha que afeta a segurança legitimamente esperada quanto ao direito de o consumidor não ser prejudicado na sua intimidade, na sua honra, na sua imagem ou no seu patrimônio.
IF = 17.53	IF = 13.095
DS = 1.043	DS = 1.000
TTR = 0.744	TTR = 0.733

Os índices da Reescrita 2 mostram que essa simplificação com acréscimo de explicação aumentou a complexidade do texto segundo o IF, mas diminuiu quanto à DS e à TTR. A diminuição do IF deve-se ao fato de terem sido acrescidas mais palavras ao trecho, variável levada em conta no cálculo desse índice. De todo modo, consideramos, neste momento, estabelecida a questão do termo *incolumidade* e, seguindo em nossa análise, verificamos que mais duas palavras do trecho não constavam no CorPop: *lide* e *versa*. O primeiro – *lide* – é um termo e significa a discussão que se resolve por meio do processo judicial. E o segundo – *versa*, de *versar* – é do léxico comum, mas não muito frequente. É empregado nos textos jurídicos com o significado de *tratar*, *abordar*. Para *lide*, colocamos como sinônimo a expressão *discussão neste processo*, e, para *versa*, colocamos a construção *é a respeito*. Como podemos ver no Quadro 6.4, a substituição se deu mais por processos sintáticos do que lexicais.

Quadro 6.4 – Reescrita 3 - Incolumidade

Reescrita 2	Reescrita 3
A lide versa a respeito de defeito do serviço, de acordo com o artigo 14, <i>caput</i> , do Código de Defesa do Consumidor, em que o objeto da relação de consumo apresenta falha que afeta a segurança legitimamente esperada quanto ao direito de o consumidor não ser prejudicado na sua intimidade, na sua honra, na sua imagem e no seu patrimônio.	A discussão neste processo é a respeito de defeito do serviço, de acordo com o artigo 14, <i>caput</i> , do Código de Defesa do Consumidor, em que o objeto da relação de consumo apresenta falha que afeta a segurança legitimamente esperada quanto ao direito de o consumidor não ser prejudicado na sua intimidade, na sua honra, na sua imagem e nos seus bens.
IF = 13.095	IF = 18.735
DS = 1.000	DS = 1.000
TTR = 0.733	TTR = 0.733

A substituição efetuada na Reescrita 3, mesmo com o acréscimo de 4 palavras, aumentou o IF, diminuindo, nesse quesito, complexidade do texto. Os índices DS e TTR mantiveram-se os mesmos.

Num terceiro momento, reduzimos algumas informações do trecho julgadas dispensáveis⁸¹ para o nosso leitor-alvo: o cidadão adulto com escolaridade máxima de até o 9º ano do Ensino Fundamental. As informações suprimidas aparecem em negrito na coluna da Reescrita 3 do Quadro 6.5, abaixo.

Quadro 6.5 – Reescrita 4 - Incolumidade

Reescrita 3	Reescrita 4
A discussão neste processo é a respeito de defeito do serviço, de acordo com o artigo 14, caput, do Código de Defesa do Consumidor, em que o objeto da relação de consumo apresenta falha que afeta a segurança legitimamente esperada quanto ao direito de o consumidor não ser prejudicado na sua intimidade, na sua honra, na sua imagem e nos seus bens.	A discussão neste processo é a respeito de defeito de serviço que afeta o direito de o consumidor não ser prejudicado na sua intimidade, na sua honra, na sua imagem e nos seus bens.
IF = 18.735	IF = 57.566
DS = 1.000	DS = 0.789
TTR = 0.733	TTR = 0.764

⁸¹ Tal supressão eliminou a referência ao texto da lei, utilizada como argumento a embasar a questão levada a análise do Judiciário, mas redundante no trecho em análise. Assim como Silva (2018, p. 125), adotamos essa estratégia com base nas orientações das *Guidelines* quanto a se posicionar no lugar do leitor-alvo e verificar o que pode ser complexo ou desnecessário para ele. Obviamente que se colocar no lugar do leitor é algo subjetivo e difícil de precisar. Justamente por isso que *corpora* como o CorPop auxiliam a verificar que palavras são mais conhecidas da população em geral.

De um trecho considerado difícil, segundo o IF, a Reescrita 4, sem as informações assinaladas na Reescrita 3, colocou o trecho entre textos considerados adequados para a escolaridade do 6º ao 9º ano.

Por fim, utilizamos de outra estratégia de simplificação: a quebra de frase, que é uma das formas de simplificação sintática. Segundo Leal (2019, p. 11), os

primeiros avanços na área de simplificação textual automática surgiram com a ideia de dividir sentenças longas em sentenças menores para melhorar os resultados dos analisadores sintáticos.

A próxima reescrita para este trecho encontra-se no Quadro 6.6, a seguir.

Quadro 6.6 – Reescrita 5 - Incolumidade

Reescrita 4	Reescrita 5
A discussão neste processo é a respeito de defeito de serviço que afeta o direito de o consumidor não ser prejudicado na sua intimidade, na sua honra e na sua imagem.	A discussão neste processo é a respeito de defeito de serviço. E o defeito de serviço afeta o direito de o consumidor não ser prejudicado na sua intimidade, na sua honra, na sua imagem e nos seus bens.
IF = 57.566	IF = 71.481
DS = 0.789	DS = 0.809
TTR = 0.764	TTR = 0.658

Para quebrar a frase, foi preciso repetir, na Reescrita 5, o termo *defeito de serviço*, representado pelo “que” da oração relativa da Reescrita 4. Essa repetição representa uma menor variação vocabular e, portanto, facilita a compreensão. Pelos índices da Reescrita 5, podemos verificar que houve uma diminuição significativa da complexidade textual em relação à Reescrita 4. Em relação ao trecho original, então, as diferenças são ainda maiores, como se vê no Quadro 6.7, a seguir:

Quadro 6.7 – Escores do Trecho Original e Reescrita 5 - Incolumidade

Trecho Original	Reescrita 5
IF = 10.330	IF = 71.481
DS = 1.136	DS = 0.809
TTR = 0,766	TTR = 0.658

O IF de 71.481 da Reescrita 5 situa o trecho como fácil e adequado para o Ensino Fundamental 2 (6º ao 9º ano), que seria o relativo ao nosso leitor-alvo. Os índices da DS e da TTR, ao diminuírem, também indicam uma menor complexidade textual.

Vejam agora esses procedimentos simplificadorios com o trecho relativo à palavra *sinistralidade*.

Sinistralidade

Esta palavra é derivada da palavra *sinistro*, que tem, entre as acepções trazidas no Dicionário Houaiss (2007), também o significado de “7 acontecimento que causa dano, perda, sofrimento ou morte; acidente, desastre; 8 prejuízo material de grande monta; dano; 9 qualquer dano havido em bem colocado no seguro”.

Com base nessas acepções, o termo *sinistralidade*, no âmbito do Direito, é entendido como a relação entre a quantidade de danos sofridos por um bem protegido por um seguro e o valor cobrado por esse seguro. Esse bem pode ser um carro, uma casa, a saúde, a vida, entre outras coisas. E o custo final de um seguro, então, vai depender da sinistralidade, ou seja, da probabilidade de ocorrência de dano ao bem que está sendo segurado. Os seguros de carro ilustram bem essa situação. Para estimar o valor do seguro de carros, a seguradora leva em conta critérios como local de residência do dono do veículo, idade do condutor etc. Por exemplo, quanto maior o índice de roubo a carros na região em que reside o proprietário do veículo, mais caro é o seguro do carro.

No contexto da Sentença onde consta este termo *sinistralidade*, tal como ocorreu com *incolumidade*, não há explicação sobre o que essa palavra significa. Assim, numa primeira proposta de simplificação, apresentamos uma reescrita em que colocamos um sinônimo para *sinistralidade*: o sintagma *previsão de dano*, que não tem ocorrência no CorPop, mas as palavras que o compõem sim: *previsão* aparece 21 vezes; e *dano*, 11. Tal reescrita é apresentada no Quadro 6.8, a seguir.

Quadro 6.8 – Reescrita 1 - Sinistralidade

Trecho Original	Reescrita 1
Aduziu que há distinção entre reajuste das mensalidades através da correção monetária, com previsão contratual, e readequação por mudança de faixa etária, que é utilizada como fator de ajuste do preço da mensalidade em razão da previsão de sua utilização com base na sinistralidade em cada faixa etária.	Aduziu que há distinção entre reajuste das mensalidades através da correção monetária, com previsão contratual, e readequação por mudança de faixa etária, que é utilizada como fator de ajuste do preço da mensalidade em razão da previsão de sua utilização com base na previsão de dano em cada faixa etária.

IF = - 4.335	IF = 1.813
DS = 1.087	DS = 1.083
TTR = 0.792	TTR = 0.760

A estratégia de simplificação lexical efetuada com a substituição de *sinistralidade* pelo sintagma *previsão de dano*, segundo os índices IF, DS e TTR apontados no Quadro 6.8, tornou a Reescrita 1 menos complexa. Chama a atenção o IF que saiu de um número negativo⁸² para um positivo.

Após, verificamos que a Reescrita 1 ainda apresentava palavras inexistentes ou pouco frequentes no CorPop. Entre as inexistentes, encontramos *aduzir*, *readequação* e *etária*. Entre as pouco frequentes, encontramos *distinção*, com 3 ocorrências; *reajuste*, com 6 ocorrências, e *ajuste*, com 11 ocorrências. Com isso, continuando num processo de simplificação lexical, apresentamos palavras mais frequentes no CorPop que pudessem atuar como sinônimo: a forma verbal *disse*, com 1342 ocorrências; *diferença*, com 48 ocorrências; *aumento*, com 33 ocorrências; *mudança*, com 74 ocorrências, e o sintagma *de idade*, com 21 ocorrências. Além dessas situações, o sintagma terminológico *correção monetária*, da área da Economia, também não consta no CorPop. Para essa situação, julgamos melhor, por um processo de simplificação por explicação, acrescentar uma explicação entre parênteses sobre o que significa essa terminologia. Tais alterações constam na Reescrita 2 no Quadro 6.9, abaixo.

Quadro 6.9 – Reescrita 2 - Sinistralidade

Reescrita 1	Reescrita 2
Aduziu que há distinção entre reajuste das mensalidades através da correção monetária , com previsão contratual, e readequação por mudança de faixa etária, que é utilizada como fator de ajuste do preço da mensalidade em razão da previsão de sua utilização com base na previsão de dano em cada faixa etária .	Disse que há diferença entre aumento das mensalidades através da correção monetária (correção do dinheiro por causa do aumento no preço das coisas), com previsão contratual, e aumento por mudança de faixa de idade, que é utilizada como fator de mudança do preço da mensalidade em razão da previsão de sua utilização com base na previsão de dano em cada faixa de idade .
IF = 1.813	IF = 7.632
DS = 1.083	DS = 0.968
TTR = 0.760	TTR = 0.635

⁸² A tabela do Índice Flesch mostra os índices a partir do zero, mas existem textos que ficam abaixo desse valor, gerando números negativos. Quanto maior o número negativo, maior a complexidade textual.

Como podemos ver pelos índices da Reescrita 2, o IF aumentou, indicando que, nesse quesito, o trecho ficou menos complexo. A TTR e a DS também diminuíram, o que significa menor complexidade também.

Após esses processos de simplificação, verificamos, a partir da Reescrita 2, outras situações que poderiam, considerando o nosso leitor-alvo, ser retiradas da frase por um processo de redução da informação⁸³. No Quadro 6.10, essas situações encontram-se grifadas:

Quadro 6.10 – Reescrita 3 - Sinistralidade

Reescrita 2	Reescrita 3
Disse que há diferença entre aumento das mensalidades através da correção monetária (correção do dinheiro por causa do aumento no preço das coisas), com previsão contratual , e aumento por mudança de faixa de idade, que é utilizada como fator de mudança do preço da mensalidade em razão da previsão de sua utilização com base na previsão de dano em cada faixa de idade.	Disse que há diferença entre aumento das mensalidades através da correção monetária (correção do dinheiro por causa do aumento no preço das coisas) e aumento por mudança de faixa de idade.
IF = 7.632	IF = 34.525
DS = 0.968	DS = 1.066
TTR = 0.635	TTR = 0.774

Como se pode ver no Quadro 6.10, com a redução das informações, a Reescrita 3 teve o seu IF aumentado consideravelmente, ainda que inclua o trecho entre os considerados difíceis. Entretanto, os dois outros índices tiveram um pequeno aumento, o que indica maior complexidade no texto. O aumento no índice da TTR pode ser explicado, porque foram retiradas informações explicativas, que tendem a repetir palavras no trecho, de modo que o texto ficou com mais palavras diferentes. No caso da DS, apesar de o índice ter aumentado, a diferença entre as palavras lexicais e funcionais ficou a mesma nos dois trechos: duas palavras lexicais a mais que as funcionais.

⁸³ Retomamos aqui a observação da nota de rodapé 18 quanto a adotar tal estratégia com base nas orientações das *Guidelines* colocando-nos no lugar do leitor-alvo para verificar o que pode ser complexo ou desnecessário para ele. No caso, as informações retiradas são redundantes.

A seguir, no Quadro 6.11, colocamos o comparativo dessa Reescrita 3 com o trecho original.

Quadro 6.11 – Escores do Trecho Original e Reescrita 3 - Sinistralidade

Trecho Original	Reescrita Final
IF = - 4.335	IF = 34.525
DS = 1.087	DS = 1.066
TTR = 0.792	TTR = 0.774

Verificamos, no Quadro 6.11, que o IF da reescrita final aumentou consideravelmente, mas ainda enquadra o trecho numa faixa de textos considerados difíceis. Em relação à DS e à TTR, houve também um decréscimo, ainda que pequeno, mas que revela igualmente uma menor complexidade textual.

Neste Estudo-Piloto 2, os dados revelaram que as sentenças e os textos do DG apresentam grande disparidade quanto ao léxico empregado. Uma das razões para isso, como observamos, pode ser o fato de as sentenças serem textos especializados e, por isso, apresentarem alta incidência de terminologia jurídica. Relativamente ao CorPop, a disparidade em relação ao léxico das sentenças foi menor, mas neste caso deve ser considerado o fato de essa lista ser quase 4 vezes maior que a lista de lemas do DG, o que gera uma probabilidade maior de abarcamento do vocabulário das sentenças. Acreditamos, porém, que num estudo maior, como é o caso da nossa pesquisa de doutoramento, essa situação não se repita, porque os *corpora* de contraste têm proporções maiores que o *corpus* de estudo – textos das sentenças. E essa proporção maior servirá para salientar o que é estatisticamente mais recorrente nas sentenças.

Relativamente à proposta de tradução intralinguística das sentenças, efetivada nas sugestões de reescritas simplificadas dos trechos em que se encontravam as palavras *incolumidade* e *sinistralidade*, ambas ausentes no CorPop, houve uma diminuição da complexidade lexical e sintática dos trechos originais. É preciso considerar que essa diminuição é relativa apenas ao Índice Flesch e ao vocabulário empregado. No que se refere à CT, muitos outros fatores devem ser considerados, que, pelos limites do Estudo-Piloto, ficaram restritos em nossas análises.

Na próxima seção, apresentamos os dados de um questionário por nós elaborado e aplicado a um grupo de leitores reais. O questionário foi desenvolvido com base nos

achados deste Estudo-Piloto 2 e pretende contemplar um dos objetivos de nossa pesquisa, conforme consta no Capítulo 3.

6.3 QUESTIONÁRIO DE VERIFICAÇÃO DA COMPLEXIDADE TEXTUAL

Nesta seção, atendendo a dois objetivos relacionados no Capítulo 3 deste trabalho, trazemos os dados de um questionário por nós elaborado e aplicado a um grupo de leitores reais. Retomamos aqui tais objetivos:

6) Propor estratégias de formulação e/ou reescrita de trechos das Sentenças considerados complexos para o nosso perfil de leitor, de modo a aproximar esses trechos aos padrões do PPE.

7) Verificar, com um grupo de leitores reais e o mais próximo possível do perfil de leitor com pouca escolaridade e experiência em leitura, se as estratégias propostas tendem a favorecer a compreensão das informações pelo público a que se destinam.

O questionário foi respondido de forma *on-line*, por sua aplicação ter coincidido com o advento da pandemia gerada pelo coronavírus responsável pela Covid-19. Em razão disso, conforme experienciamos, também ficou prejudicado o acesso a mais respondentes que preenchessem, justamente, o perfil de leitor que temos em mente neste trabalho: um cidadão brasileiro, adulto com escolaridade máxima de até o 9º ano do Ensino Fundamental. O encaminhamento do nosso questionário aos respondentes se deu por meio de contatos de pessoas amigas, que, por sua vez, tinham acesso a outras pessoas que poderiam, eventualmente, preencher o perfil desejado.

O questionário foi aplicado considerando o parecer do Comitê de Ética da UFRGS aprovado: CAAE: 59792716.8.0000.5347, que foi exarado para o Projeto de Pesquisa *Da Doença de Parkinson a cuidados básicos em Pediatria, acessibilidade textual e terminológica para leitores brasileiros de baixa escolaridade*. Tal projeto insere-se no âmbito das pesquisas sobre Acessibilidade Textual e Terminológica coordenadas pela Professora. Maria José Finatto. E, na aplicação dos questionários propostos, estão previstos também outros tipos e temas de textos, que servem como contraponto. É o caso da nossa pesquisa, que, mesmo não sendo relativa à área da saúde, trata da acessibilidade textual e terminológica também.

A ferramenta utilizada para criação e distribuição do questionário foi a SurveyMonkey⁸⁴, plataforma de serviços *on-line*, fundada por Ryan Finley em 1999:

SurveyMonkey provê pesquisas personalizáveis gratuitas, bem como uma suíte de programas *back-end* que inclui análise de dados, seleção de amostras, eliminação de vieses, e ferramentas de representação de dados.⁸⁵

No Estudo-Piloto 2, após a comparação do léxico de um grupo de Sentenças dos JECs com palavras do CorPop e do jornal Diário Gaúcho, reescrevemos, utilizando várias estratégias simplificadoras, alguns trechos que continham palavras consideradas “complexas”⁸⁶. Assim, para a elaboração do questionário envolvendo nossas simplificações, nos servimos de alguns exemplos dessas palavras complexas e de algumas reescritas propostas para os trechos em que elas se encontravam.

O questionário que aplicamos foi organizado da seguinte forma:

- a) apresentação da autora do questionário e das razões para a sua aplicação;
- b) três questões para traçar o perfil dos respondentes quanto à idade, escolaridade e renda familiar;
- c) cinco questões envolvendo o conhecimento do significado das palavras;
- d) quatro questões envolvendo a compreensão com base em reescritas de alguns trechos;
- e) agradecimento pela realização do questionário.

Antes de ser encaminhado aos respondentes, o questionário foi enviado a alguns colegas⁸⁷ do nosso grupo de Pesquisa em ATT. As sugestões trazidas por esses colegas ajudaram a aperfeiçoar as questões e a chegar à versão final do instrumento que foi aplicado.

A seguir colocamos a íntegra do instrumento, tal como disponibilizado na plataforma SurveyMonkey.

⁸⁴ **SurveyMonkey**. Disponível em: <https://pt.surveymonkey.com/>

⁸⁵ **SurveyMonkey**. In: Wikipédia. Disponível em: <https://pt.wikipedia.org/wiki/SurveyMonkey>

⁸⁶ Foram consideradas “complexas” porque não constavam no CorPop, por este *corpus* ser um *corpus* de referência do PPE que pode servir como baliza para a estimativa da complexidade lexical de um texto.

⁸⁷ Bruna Rodrigues da Silva, Eduardo Felipe Felten, Liana Braga Paraguassu, Lucas Meireles Tcacenco e Gabriel Luciano Ponomarenko.

Página 1 do Questionário

Olá, prezado respondente:

Sou aluna do Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul e minha pesquisa é sobre a linguagem empregada em sentenças judiciais e como ela é compreendida pela população em geral.

O título de minha tese em andamento é *Sentenças Judiciais e Linguagem Simples: um encontro possível e necessário*.

Uma das etapas da minha pesquisa é este questionário encaminhado a você. Suas respostas são muito importantes para o trabalho. Por isso, deixo desde já o meu agradecimento por sua participação.

Ester Motta.
(contato: estermottac@gmail.com)

IMPORTANTE:

Seu nome não fica registrado no questionário.

Não é preciso nenhum conhecimento anterior sobre a linguagem jurídica. A ideia é saber como as pessoas em geral compreendem a linguagem do Direito.

São apenas 12 perguntas de marcar, mas todas precisam ser respondidas.

Páginas 2 e 3 do Questionário – Questões 1 a 3

*** 1. Antes de tudo, eu gostaria de conhecer um pouquinho sobre você.**

Não se preocupe, pois **seu nome não fica registrado no questionário.**

Qual é a sua idade?

- Entre 16 e 21 anos.
- Entre 22 e 30 anos.
- Entre 30 e 40 anos.
- Entre 40 e 50 anos.
- Entre 50 e 60 anos.
- Mais de 61 anos.

*** 2. Você estudou (em escolas ou outros locais) até:**

- Ensino Fundamental Incompleto
- Ensino Fundamental Completo (até o 8º ou o 9º ano)
- Ensino Médio Incompleto
- Ensino Médio Completo
- Ensino Superior (ainda está cursando)
- Ensino Superior Completo
- Ensino de Pós-Graduação

*** 3. A renda da sua família é de:**

- de 1 a 3 salários mínimos
- de 3 a 5 salários mínimos
- de 5 a 7 salários mínimos
- de 7 a 10 salários mínimos
- mais de 10 salários mínimos

Página 4 do Questionário – Questões 4 a 8

*** 4. O TRECHO 1, a seguir, foi retirado de uma sentença dos Juizados de Pequenas Causas (Juizados Especiais Cíveis do Poder Judiciário do Estado do Rio Grande do Sul). Uma sentença judicial é o texto em que o Juiz diz o que decidiu sobre um problema levado ao Judiciário.**

Peco que você leia o TRECHO 1.

TRECHO 1

Aduziu que há distinção entre reajuste das mensalidades através da correção monetária, com previsão contratual, e readequação por mudança de faixa etária, que é utilizada como fator de ajuste do preço da mensalidade em razão da previsão de sua utilização com base na sinistralidade em cada faixa etária.

Agora, marque, em cada item abaixo, o que você acha que significa, no Trecho 1, cada palavra destacada.

4. ADUZIU:

- significou
- disse
- reclamou
- questionou

*** 7. SINISTRALIDADE:**

- previsão de choque
- previsão de dano
- previsão de discórdia
- previsão de contrato

*** 5. DISTINÇÃO:**

- diferença
- semelhança
- relação
- confusão

*** 8. FAIXA ETÁRIA:**

- faixa de doença
- faixa de temperatura
- faixa de idade
- faixa de dinheiro

*** 6. READEQUACÃO:**

- reconforto
- reajuste
- revalidação
- recomendação

Página 5 do Questionário – Questão 9

*** 9. Podemos dizer a mesma coisa de diferentes maneiras. É o que acontece nas frases (1a) e (1b) abaixo:**

(1a) O juiz decidiu o caso.

(1b) O caso foi decidido pelo juiz.

As duas frases são diferentes, mas trazem a mesma informação: que **alguém (o Juiz) decidiu alguma coisa (o caso).**

Pensando nessa possibilidade de dizer a mesma coisa de jeitos diferentes, marque a alternativa que MELHOR PASSA A IDEIA do Trecho 1, que está repetido abaixo.

TRECHO 1

Aduziu que há distinção entre reajuste das mensalidades através da correção monetária, com previsão contratual, e readequação por mudança de faixa etária, que é utilizada como fator de ajuste do preço da mensalidade em razão da previsão de sua utilização com base na sinistralidade em cada faixa etária.

- Aduziu que a distinção entre reajuste das mensalidades pela correção monetária e pela idade é mínima. A mudança pela correção monetária é feita no contrato e a mudança pela idade é feita pensando no quanto a pessoa vai utilizar o seguro conforme a sua faixa etária.
- Afirmou que o reajuste das mensalidades pela correção monetária e pela idade é diferente. O reajuste pela correção monetária é estabelecido no contrato, e o reajuste pela idade é feito com base no quanto a pessoa poderá utilizar o seguro conforme a sua faixa etária.
- A distinção entre o reajuste das mensalidades pela correção monetária e pela idade aduzida acontece porque a mudança pela correção monetária é estabelecida no contrato, e a mudança pela idade não.

Página 6 do Questionário – Questões 10 a 12

10. Entre os muitos jeitos de se escrever alguma coisa, alguns são mais fáceis, uns não são nem fáceis nem difíceis, e outros são muito difíceis. Claro que o “fácil” e o “difícil” vai depender de cada pessoa. Mas vamos imaginar que três políticos tenham de falar ao povo brasileiro na televisão e se manifestem da seguinte forma:

Político A:

Esforçar-nos-emos para viabilizar uma sociedade mais inclusiva.

Político B:

Vamos nos esforçar para viabilizar uma sociedade mais inclusiva.

Político C:

Queremos construir um país em que ninguém fique de fora.

Qual político usou um jeito de falar mais próximo do povo brasileiro?

Imagino que você tenha respondido que **foi o Político C.**

Pensando que 1 seria uma nota para a fala mais difícil, 2 uma nota para a mais ou menos, e 3 uma nota para a mais fácil, as falas acima teriam a seguinte classificação:

Político A: 1 (fala mais difícil)

Político B: 2 (fala mais ou menos difícil)

Político C: 3 (fala mais fácil)

Agora vou pedir que você, pensando nos exemplos dos políticos acima:

- a) considere essa possibilidade de dizer/escrever coisas de forma mais fácil ou mais difícil,**
- b) leia as Escritas 1, 2 e 3, escritas diferentes de um mesmo trecho; e**
- c) dê uma nota/classificação a elas, pensando da seguinte forma:**

1 – para a que você achar MAIS DIFÍCIL de entender o que está escrito;

2 - para a que você achar MAIS OU MENOS DIFÍCIL de entender o que está escrito;

3 - para a que você achar a MAIS FÁCIL de entender o que está escrito.

Atenção:

Não pode haver repetição da classificação. Ou seja: não pode haver duas escritas com a mesma classificação.

ESCRITA 1

A lide versa a respeito de defeito do serviço, de acordo com artigo 14, caput, do Código de Defesa do Consumidor, em que o objeto da relação de consumo

apresenta falha que afeta a segurança legitimamente esperada quanto à incolumidade subjetiva do consumidor ou do seu patrimônio.

- 1 (MAIS DIFÍCIL de entender o que está escrito)
- 2 (MAIS OU MENOS DIFÍCIL de entender o que está escrito)
- 3 (MAIS FÁCIL de entender o que está escrito)

*** 11. ESCRITA 2**

A discussão neste processo é a respeito de defeito de serviço. O art. 14 do Código de Defesa do Consumidor diz que o defeito de serviço afeta o direito de o consumidor não ser prejudicado na sua intimidade, na sua honra, na sua imagem e nos seus bens.

- 1 (MAIS DIFÍCIL de entender o que está escrito)
- 2 (MAIS OU MENOS DIFÍCIL de entender o que está escrito)
- 3 (MAIS FÁCIL de entender o que está escrito)

*

12. ESCRITA 3

A discussão neste processo é a respeito de defeito do serviço, de acordo com o artigo 14, caput, do Código de Defesa do Consumidor, em que o objeto da relação de consumo apresenta falha que afeta a segurança legitimamente esperada quanto ao direito de o consumidor não ser prejudicado na sua intimidade, na sua honra, na sua imagem e nos seus bens.

- 1 (MAIS DIFÍCIL de entender o que está escrito)
- 2 (MAIS OU MENOS DIFÍCIL de entender o que está escrito)
- 3 (MAIS FÁCIL de entender o que está escrito)

O questionário termina aqui.

**Quando minha pesquisa finalizar, ela e os dados deste questionário estarão disponíveis para acesso livre e gratuito na página <https://www.lume.ufrgs.br/>.
Muito obrigada pela sua VALIOSA COLABORAÇÃO**

Tendo apresentado o instrumento aplicado, passamos, na próxima seção, à análise das respostas recolhidas.

6.3.1 Análise dos Dados

As primeiras questões do questionário foram respondidas por 99 respondentes. As demais tiveram uma diminuição gradual nesse número de respondentes. Ao finalizarmos as análises das respostas, fazemos algumas observações sobre este fato.

Inicialmente, então, para identificarmos o perfil dos respondentes, solicitamos informações quanto à idade, escolaridade e renda familiar. Como nosso trabalho tem como protótipo de leitor-alvo um cidadão brasileiro, adulto com escolaridade máxima de até o 9º ano do Ensino Fundamental⁸⁸, julgamos importante esses dados para verificar o quanto os nossos respondentes se encaixariam nesse perfil.

A Tabela 6.8, a seguir, apresenta os dados relativos à idade dos respondentes.

Tabela 6.8 – Respondentes por Faixa Etária

FAIXA ETÁRIA	QUANTIDADE	PERCENTUAL
Entre 16 e 21 anos	30	30,30%
Entre 22 e 30 anos	13	13,13%
Entre 30 e 40 anos	9	9,09%
Entre 40 e 50 anos	14	14,14%
Entre 50 e 60 anos	19	19,19%
Mais de 61 anos	14	14,14%
TOTAL	99	100%

Pela Tabela 6.8, verificamos que a maior parte dos respondentes têm entre 16 e 21 anos, seguidos daqueles que estão entre 50 e 60 anos. Porém, se somarmos os respondentes que contam 22 anos para cima – que são, em princípio, os cidadãos economicamente ativos –, temos um total de 66 respondentes que se aproximam, nesse quesito, do perfil de leitor imaginado em nossa pesquisa.

A seguir, na Tabela 6.9, temos os dados relativos à escolaridade dos respondentes.

Tabela 6.9 – Respondentes por Escolaridade

ESCOLARIDADE	QUANTIDADE	PERCENTUAL
Ensino Fundamental Incompleto	3	3,10%
Ensino Fundamental Completo (até o 8º ou 9º ano)	9	9,20%
Ensino Médio Incompleto	17	17,34%

⁸⁸ Conforme já mencionamos nesta pesquisa, são leitores das classes C, D e E, com escolaridade correspondente ao Ensino Fundamental completo e pouca experiência em leitura.

Ensino Médio Completo	21	21,40%
Ensino Superior (ainda está cursando)	18	18,36%
Ensino Superior Completo	19	19,38%
Ensino de Pós-Graduação	11	11,22%
TOTAL DE RESPONDENTES	98	100%

Segundo a Tabela 6.9, a maioria dos nossos respondentes tem escolaridade a partir do Ensino Médio. Entre essa faixa, a maior parte tem Ensino Médio Completo, seguidos dos que têm Ensino Superior Completo. Ainda que essas características não contemplem o perfil de leitor-alvo imaginado em nossa pesquisa, é um dado a ser considerado na totalidade das respostas ao questionário. Como já comentamos no decorrer do nosso trabalho, inclusive com uma citação do escritor Luís Fernando Veríssimo, a queixa em relação à dificuldade de se compreender a linguagem jurídica parece ser geral, independentemente de escolaridade.

Essa situação é confirmada na pesquisa *O Estudo da Imagem do Poder Judiciário Brasileiro*, realizada pela Fundação Getúlio Vargas (FGV), em parceria com o Instituto de Pesquisas Sociais, Políticas e Econômicas (IPESPE), por nós mencionada na seção 1.1. Segundo essa pesquisa, as pessoas com mais escolaridade consideram a linguagem jurídica pouco compreensível em maior proporção que as com menos escolaridade. Trata-se de diferenças mínimas – Ensino Fundamental (85%), Ensino Médio (88%), Ensino Superior (89%) -, mas é um dado que, em princípio, confere com o fato de ser geral a queixa em relação à falta de transparência da linguagem jurídica.

Além disso, cumpre referir aqui a pesquisa de doutorado realizada por Bianca Pasqualini, que deu origem ao CorPop. Ela cita o Jornal Boca de Rua, de Porto Alegre e redigido por pessoas em situação de vulnerabilidade social. Pasqualini comenta que muitos se alfabetizam para poder participar da redação do jornal, e os textos que escrevem apresentam uma sintaxe e um léxico de certo modo complexos. Com base nessa experiência, ela afirma: “Assim, podemos inferir que a prática de leitura tem correlação direta com uma melhora progressiva do letramento do leitor, mas não necessariamente com seu nível de escolaridade” (PASQUALINI, 2018, p. 133).

Para corroborar essa afirmação de Pasqualini quanto a essa relação entre letramento e escolaridade, retomamos aqui os dados do INAF-2018: entre as pessoas que completaram o Ensino Médio, apenas 12% demonstraram domínio pleno das habilidades

de leitura, escrita e matemática. Já quanto ao Ensino Superior, completo ou incompleto, apenas 34% das pessoas encontram-se no nível pleno de alfabetismo.

E, por fim, em relação ao perfil dos nossos respondentes, temos a Tabela 6.10, que se refere à renda familiar.

Tabela 6.10 – Respondentes por Renda Familiar

RENDA FAMILIAR	QUANTIDADE	PERCENTUAL
De 1 a 3 salários mínimos	43	44,79%
De 3 a 5 salários mínimos	17	17,71%
De 5 a 7 salários mínimos	7	7,29%
De 7 a 10 salários mínimos	8	8,33%
Mais de 10 salários mínimos	21	21,88%
TOTAL	96	100%

Segundo a Tabela 6.10, as duas maiores faixas são as opostas, ou seja, 43 dos 96 respondentes que afirmam ter renda familiar de 1 a 3 salários mínimos, enquanto 21 afirmam ter renda superior a 10 salários mínimos. A informação sobre a renda familiar nos permite inferir sobre o acesso que, em princípio, esses respondentes têm a materiais escritos.

Estando a maioria na faixa que vai de 1 a 3 salários-mínimos, podemos inferir que, em tese, o acesso a materiais escritos, que implica acesso à informação e à cultura, é menor que o que pode se dar nas demais faixas. Isso, por sua vez, pode influenciar no nível de proficiência desses leitores. Segundo Pasqualini (2018, p. 16),

[m]esmo com a internet popularizada, com a quantidade de textos e informações disponíveis, vemos que a maioria da população carece de uma formação leitora suficiente para lidar com esse todo. Mesmo exposta à informação, mesmo exposta a textos, a livros, seja na internet ou em suporte físico, apenas uma pequena parcela consegue interagir qualificadamente com a informação que recebe.

Esses são os dados relacionados ao perfil do nosso respondente. Com eles podemos, em termos gerais, dizer que a maioria dos nossos respondentes está na faixa dos 16 a 21 anos, com o Ensino Médio Completo e renda familiar de até 3 salários mínimos no máximo.

A partir da Questão 4, iniciamos os questionamentos de conteúdo. Como já mencionamos, as questões de n. 4 a 8 foram relativas ao significado das palavras. O trecho que deu origem a essas questões foi escolhido por ser o contexto em que a palavra

sinistralidade se encontrava. *Sinistralidade* foi uma das palavras que não constavam no CorPop – utilizado nas comparações realizadas no Estudo-Piloto 2. Além disso, também não há, no contexto da sentença de onde ela foi extraída, alguma forma de explicação (uma paráfrase, por exemplo) para o que ela significa, tal como explicamos no Estudo-Piloto 2.

Como podemos observar, além da palavra *sinistralidade*, no trecho escolhido foi possível selecionar outras três palavras e/ou expressões – *aduziu*, *readequação* e *faixa etária* – também ausentes no CorPop. E *distinção* consta com apenas 3 ocorrências. Diante disso, uma vez que o CorPop é um *corpus* que representa o português popular escrito do Brasil, em princípio, as palavras assinaladas no trecho não são conhecidas do leitor brasileiro médio.

Porém, devemos considerar que essas 5 palavras se encontram inseridas num trecho maior, que, segundo os estudos sobre a compreensão leitora, pode ser um facilitador na identificação do significado da leitura. É o que muitos chamam de mecanismo *top-down*, que ocorre quando o leitor se apoia no contexto ou no seu conhecimento prévio para chegar ao significado de alguma palavra desconhecida.

Observamos também que essas 5 palavras assinaladas no trecho em questão representam 10,41% do total de 48 palavras constantes no trecho, segundo a contagem do editor de texto *Word*. No AntConC, o trecho ficou com 48 *tokens* e 38 *types*. Nesse caso, o percentual das 5 palavras em relação aos *types* ficaria 13,16%. Fulgêncio e Liberato (2010, p. 105) trazem dados de pesquisas quanto a essa situação relativa à quantidade de palavras desconhecidas em um texto:

Estudos de Hu e Nation (*apud* Nation, 2001: 147) evidenciaram uma nítida relação entre compreensão de texto e de vocabulário, verificando que a compreensão do texto melhora na medida em que aumenta o número de palavras conhecidas. Se o texto apresenta 95% de palavras conhecidas, alguns leitores compreendem o texto, mas muitos não adquirem adequada compreensão; se a porcentagem é de 90% de palavras conhecidas, somente um pequeno grupo de pessoas consegue compreender o texto; e no nível de 80% de vocabulário conhecido, nenhum leitor consegue compreender o texto. Hu e Nation concluíram que, para garantir a compreensão, um texto deve conter por volta de 98% de palavras conhecidas. Carver (1994: 432, *apud* Nation, 2001: 148) acrescenta ainda que um texto apropriado, ou seja, não considerado difícil, contém somente 1% ou menos de palavras desconhecidas.

Além das 5 palavras assinaladas, *contratual* e *correção monetária* igualmente não constam no CorPop. Com isso e segundo os dados das pesquisas da citação acima, o trecho de nossa primeira questão seria compreendido apenas por um pequeno grupo de pessoas.

O que essas pesquisas relatam confirmam o que muitos estudiosos da compreensão leitora comentam em relação ao mecanismo *top-down*. A presença de muitas palavras desconhecidas impede o leitor de prever aquela palavra e as outras que a acompanham, não podendo montar uma representação mental do texto. Fulgêncio e Liberato (2010, p. 108) comentam que não só palavras desconhecidas prejudicam esse processo, mas também o emprego de palavras raras

é prejudicial à legibilidade, na medida em que provoca uma duração de fixação mais longa do que em palavras frequentes, atrasando a fluência da leitura. Isso reflete a maior demanda cognitiva para o acesso lexical, isto é, uma maior dificuldade para encontrar na memória a representação mental da palavra.

Agora vejamos como foram as respostas de nossos respondentes no que diz respeito a essas questões. As Tabelas 6.11 a 6.15 mostram o escore de cada questão.

Tabela 6.11 – Escore da Questão 4 - Aduziu

ALTERNATIVAS	RESPOSTAS	PERCENTUAL
significou	14	16,87%
disse	44	53,01%
reclamou	6	7,23%
questionou	19	22,89%
TOTAL	83	100%

Tabela 6.12 – Escore da Questão 5 - Distinção

ALTERNATIVAS	RESPOSTAS	PERCENTUAL
diferença	74	89,16%
semelhança	3	3,61%
relação	5	6,02%
confusão	1	1,20%
TOTAL	83	100%

Tabela 6.13 – Escore da Questão 6 - Readequação

ALTERNATIVAS	RESPOSTAS	PERCENTUAL
reconforto	4	4,82%
reajuste	71	85,54%
revalidação	6	7,23%
recomendação	2	2,41%
TOTAL	83	100%

Tabela 6.14 – Escore da Questão 7 - Sinistralidade

ALTERNATIVAS	RESPOSTAS	PERCENTUAL
previsão de choque	6	7,23%
previsão de dano	47	56,63%
previsão de discórdia	5	6,02%
previsão de contrato	25	30,12%
TOTAL	83	100%

Tabela 6.15 – Escore da Questão 8 - Faixa Etária

ALTERNATIVAS	RESPOSTAS	PERCENTUAL
faixa de doença	1	1,20%
faixa de temperatura	0	0%
faixa de idade	80	96,39%
faixa de dinheiro	2	2,41%
TOTAL	83	100,00%

E a Tabela 6.16 apresenta o percentual de respostas corretas para cada questão em ordem decrescente.

Tabela 6.16 – Respostas Corretas das Questões 4 a 8 em Ordem Decrescente

QUESTÃO	RESPOSTA CORRETA	QUANTIDADE DE RESPOSTAS	PERCENTUAL DE RESPOSTAS
8 - Faixa etária	Faixa de idade	80	96,39%
5 - Distinção	Diferença	74	89,16%
6 - Readequação	Reajuste	71	85,54%
7 - Sinistralidade	Previsão de Dano	47	56,63%
4 - Aduziu	Disse	44	53,01%

Chamamos a atenção para a redução gradual no número de respondentes. Iniciaram a responder 99 pessoas, após ficaram 96, e para essas questões de conteúdo, ficaram 83 respondentes.

Entre as questões relativas ao significado das palavras, a que ofereceu maior dificuldade aos respondentes foi *aduziu*. Uma das razões para esse resultado pode ser o fato de esta ser a primeira palavra do trecho. Nesse caso, o leitor ainda precisa prosseguir

na leitura para descobrir pistas que o ajudem a construir o mapa mental do texto e pelo menos ter uma ideia do que tal palavra significa. Além disso, três das alternativas propostas são muito semelhantes em termos de significado: *disse* – *reclamou* e *questionou*. Essas três palavras fazem parte de um mesmo esquema de conceitos relacionados a *atos de proferir algo*. A que destoa um pouco é *significou*, cujo significado não tem muita relação com as outras opções. Ainda assim, um percentual de 16,87 de respostas marcaram esta opção.

Em segundo lugar, em termos de dificuldade, ficou a questão relativa à *sinistralidade*, com pouca diferença de acerto em relação a *aduziu*, apenas 3 respondentes a mais. *Sinistralidade*, como vimos, foi a que nos levou a escolher o trecho da questão. É interessante notar que, entre as respostas dadas nesta questão, a que ficou em segundo lugar foi *previsão de contrato*. Os respondentes talvez tenham escolhido esta opção por haver nela a palavra *contrato*, muito utilizada nos documentos jurídicos. Assim, pode ser que, mesmo não sabendo o exato significado da palavra *sinistralidade*, imaginaram que, pelo trecho ser de uma sentença judicial e havendo uma opção com *contrato*, um termo jurídico, esta seria uma opção possível.

As outras palavras, conforme vemos na Tabela 6.16, tiveram um percentual de acerto bem alto, acima de 80%, o que nos indica que, mesmo não constando no CorPop, não foram de difícil compreensão. Cumpre referir que a maioria dos nossos respondentes tem o Ensino Médio Completo e, por essa razão, podem ter mais contato com materiais escritos de diversas ordens. Entretanto, podemos inferir que, para alguém que tenha apenas o Ensino Fundamental e tenha pouca experiência em leitura, palavras como *aduziu* e *sinistralidade* sejam ainda mais desconhecidas e, portanto, complexas.

As Questões 9 a 12 envolvem questionamentos sobre a compreensão em diferentes reescritas de um mesmo texto. O trecho da Questão 9 foi o mesmo das Questões 4 a 8, e a proposta era verificar qual reescrita apresentada nas três alternativas diria o mesmo que o trecho original, repetido a seguir.

Aduziu que há distinção entre reajuste das mensalidades através da correção monetária, com previsão contratual, e readequação por mudança de faixa etária, que é utilizada como fator de ajuste do preço da mensalidade em razão da previsão de sua utilização com base na sinistralidade em cada faixa etária.

As respostas dadas com suas respectivas quantidades encontram-se na Tabela 6.17, a seguir.

Tabela 6.17 – Respostas Dadas à Questão 9

ALTERNATIVAS	RESPOSTAS	PERCENTUAL
Aduziu que a distinção entre reajuste das mensalidades pela correção monetária e pela idade é mínima. A mudança pela correção monetária é feita no contrato e a mudança pela idade é feita pensando no quanto a pessoa vai utilizar o seguro conforme a sua faixa etária.	12	15,79%
Afirmou que o reajuste das mensalidades pela correção monetária e pela idade é diferente. O reajuste pela correção monetária é estabelecido no contrato, e o reajuste pela idade é feito com base no quanto a pessoa poderá utilizar o seguro conforme a sua faixa etária.	56	73,68%
A distinção entre o reajuste das mensalidades pela correção monetária e pela idade aduzida acontece porque a mudança pela correção monetária é estabelecida no contrato, e a mudança pela idade não.	8	10,53%
TOTAL	76	100%

Dos 83 que responderam às 8 primeiras questões, apenas 76 responderam a esta.

Em relação à questão propriamente dita, diferentemente das anteriores, que só trabalharam com o léxico, ela inclui uma reformulação sintática do trecho. Inclusive, com a realização dessa questão, o respondente poderia encontrar pistas para as respostas às questões anteriores e eventualmente retornar a elas para corrigir a resposta dada. Uma das pistas poderia ser a troca do verbo *aduziu* por *afirmou*. Ainda que o respondente tivesse dúvidas quanto à correção da segunda opção desta questão, poderia estabelecer uma relação do verbo *afirmou* com *disse*, que tem significado próximo.

As duas primeiras alternativas transformam o trecho de uma frase só em duas. A primeira mantém o vocabulário utilizado no trecho original – *aduziu*, *distinção*, *reajuste* –, só colocando a palavra *idade*, em substituição à *etária*, na primeira frase. Porém, muda a mensagem do trecho original, que afirma, com a forma verbal *há*, existir uma distinção, uma diferença, entre os tipos de reajustes. A reescrita afirma que essa distinção é mínima, trazendo uma ideia que não existe no trecho original. Mesmo assim, houve 12 respostas para essa alternativa.

Na segunda alternativa, o verbo *aduziu* foi substituído por *afirmou* e estabelece uma relação entre reajuste pela correção monetária e reajuste pela idade. Depois, na segunda frase, em vez de usar *contratual*, utilizou *estabelecido no contrato*. E a terceira alternativa estabelece uma relação de causalidade que não existe no trecho original.

O percentual de “acertos” da questão foi bem alto: dos 76 respondentes, a grande maioria, 73,68%, assinalou a resposta correta.

Nas Questões 10, 11 e 12, apresentamos duas reescritas a outro trecho que simplificamos no Estudo-Piloto 2: a original, que a é Escrita 1, e mais duas reescritas – Escrita 2 e Escrita 3, uma das versões simplificadas no Estudo-Piloto 2. A solicitação foi a apreciação em termos de grau de facilidade para a compreensão.

As Tabelas 6.18, 6.19 e 6.20 mostram os trechos trabalhados e as respectivas respostas.

Tabela 6.18 – Respostas Dadas à Questão 10 – Escrita

10 – ESCRITA 1		
<i>A lide versa a respeito de defeito do serviço, de acordo com artigo 14, caput, do Código de Defesa do Consumidor, em que o objeto da relação de consumo apresenta falha que afeta a segurança legitimamente esperada quanto à incolumidade subjetiva do consumidor ou do seu patrimônio.</i>		
Opções de resposta	Respostas	Percentual
1 (MAIS DIFÍCIL de entender o que está escrito)	54	81,82%
2 (MAIS OU MENOS DIFÍCIL de entender o que está escrito)	5	7,58%
3 (MAIS FÁCIL de entender o que está escrito)	7	10,60%
TOTAL	66	100%

Tabela 6.19 – Respostas Dadas à Questão 11 – Escrita 2

11– ESCRITA 2		
<i>A discussão neste processo é a respeito de defeito de serviço. O art. 14 do Código de Defesa do Consumidor diz que o defeito de serviço afeta o direito de o consumidor não ser prejudicado na sua intimidade, na sua honra, na sua imagem e nos seus bens.</i>		
Opções de resposta	Respostas	Percentual
1 (MAIS DIFÍCIL de entender o que está escrito)	3	4,55%
2 (MAIS OU MENOS DIFÍCIL de entender o que está escrito)	17	25,75%
3 (MAIS FÁCIL de entender o que está escrito)	46	69,70%
TOTAL	66	100%

Tabela 6.20 – Respostas Dadas à Questão 12 – Escrita 3**12– ESCRITA 3**

A discussão neste processo é a respeito de defeito do serviço, de acordo com o artigo 14, caput, do Código de Defesa do Consumidor, em que o objeto da relação de consumo apresenta falha que afeta a segurança legitimamente esperada quanto ao direito de o consumidor não ser prejudicado na sua intimidade, na sua honra, na sua imagem e nos seus bens.

Opções de resposta	Respostas	Percentual
1 (MAIS DIFÍCIL de entender o que está escrito)	3	4,55%
2 (MAIS OU MENOS DIFÍCIL de entender o que está escrito)	42	63,63%
3 (MAIS FÁCIL de entender o que está escrito)	21	31,82%
TOTAL	66	100%

Com os dados das três tabelas anteriores, geramos a Tabela 6.21, colocando os trechos por ordem de dificuldade, conforme as respostas dadas ao questionário.

Tabela 6.21 – Respostas das Questões 10, 11 e 12 por Ordem de Dificuldade

MAIS DIFÍCIL		MAIS OU MENOS DIFÍCIL		MENOS DIFÍCIL	
Questão	Respostas	Questão	Respostas	Questão	Respostas
10 (Escrita 1)	54	12 (Escrita 3)	42	11 (Escrita 2)	46
11 (Escrita 2)	3	11 (Escrita 2)	17	12 (Escrita 3)	21
12 (Escrita 3)	3	10 (Escrita 1)	5	10 (Escrita 1)	7

Inicialmente, observamos novamente uma redução nos respondentes: dos 76 que responderam até a Questão 9, apenas 66 responderam a essas últimas questões

Em relação às respostas, o texto considerado mais difícil pelos respondentes foi o da Escrita 1, que é o trecho original, extraído de uma das sentenças do nosso CE. Os dados do Estudo-Piloto 2 analisados para este trecho o qualificam como *muito difícil*. São eles: IF = 10.330, DS = 1.136 e TTR = 0,766.

O IF, que se refere ao Índice Flesch, considera a extensão da frase e das palavras. No capítulo em que tratamos da ATT, colocamos uma tabela de facilidade na leitura baseada no IF. Segundo essa tabela, a Escrita 1 é um texto muito difícil e requer no mínimo o Ensino Superior para ser compreendido. A DS, que se refere à Densidade Semântica, como explicamos no Estudo-Piloto 2, mede a relação entre palavras lexicais e palavras funcionais. Quanto maior o seu índice, mais palavras lexicais apresenta o texto

e, conseqüentemente, mais palavras o leitor deve conhecer. E a TTR é a relação entre palavras diferentes (*types*) e o total de palavras (*tokens*) do texto. Como já mencionamos nesta pesquisa, quanto maior o número dessa relação, mais variado é o léxico do texto.

Os dados trazidos no Estudo-Piloto 2 conferem com a percepção de leitores reais. E este é um indicativo importante nesta pesquisa, porque foi exatamente este o objetivo do questionário: verificar o quanto a percepção de leitores reais confirma a avaliação feita pelas ferramentas computacionais utilizadas em nosso trabalho.

A Escrita 3 (Questão 12), em segundo lugar em termos de dificuldade, foi uma das reescritas simplificadas por nós no Estudo-Piloto 2. Essa reescrita ficou com os seguintes índices: IF = 18.735, DS = 1.000 e TTR = 0.733. Os índices revelam uma menor complexidade. Mesmo assim, essa Escrita 3 ainda pode ser considerada um trecho de difícil compreensão. E essa foi, em princípio, a impressão dos respondentes.

E o trecho considerado mais fácil pelos respondentes foi a Escrita 2 (Questão 11). Esta não foi extraída do Estudo-Piloto 2, mas criada para o questionário. Utilizando os mesmos índices do Estudo-Piloto 2, a Escrita 2 ficou com a seguinte pontuação: IF = 77.495; DS = 1,086; TTR = 0,667. Pelo IF já se percebe uma grande diferença – de 10.330 do trecho original para 77.495. Essa pontuação coloca o trecho como de muito fácil compreensão, bastando ter o Ensino Fundamental para compreendê-lo. A TTR ficou um pouco mais baixa que as outras reescritas, o que também é um facilitador, porque o texto ficou menos variado lexicalmente, ou seja, há menos palavras a serem conhecidas do leitor. A DS ficou um pouco maior que a da Escrita 3, o que seria um indicativo de maior dificuldade, porque revela maior presença de palavras de conteúdo no texto. Mas a diferença não é tão expressiva: de 1000 para 1086. Mais uma vez a avaliação de um leitor real confirma os dados extraídos das ferramentas computacionais utilizadas neste trabalho.

Um dado interessante a ser observado é que essa Escrita 2 é a única do grupo que apresenta 2 frases. E ela foi considerada a mais fácil, confirmando o resultado de várias pesquisas, como a do Projeto PorSimples e de Sidney Leal (LEAL, 2019), já mencionadas ao longo de nosso texto. Ou seja, a quebra de frases é um facilitador da leitura e, conseqüentemente, da compreensão.

Antes de encerrarmos, julgamos importante retomar nossas observações sobre a redução gradual no número de respondentes à medida que o questionário avançava. O número de respondentes iniciou com 99 e terminou com 66. Essa redução pode ser atribuída ao fato de que responder a um questionário, mesmo que de forma anônima e

com apenas um clique, sem ter um maior comprometimento com o tema, pode ser algo desagradável. Além disso, as questões de conteúdo exigem, logicamente, uma maior atenção. E as Questões 9, 10, 11 e 12, por envolverem diferentes escritas, demandam uma leitura mais atenta e, conseqüentemente, mais tempo. Talvez isso também tenha influenciado os desistentes.

Além dessas possibilidades, há também o fato de o questionário todo ser construído com trechos extraídos de uma sentença judicial que, por ser um texto especializado, apresenta uma linguagem também especializada que nem sempre é familiar aos respondentes. Aliás, o foco de nossa pesquisa é este: verificar o quanto a linguagem empregada nas Sentenças dos JECs é acessível, ou não, ao cidadão brasileiro com pouca experiência em leitura e escolaridade limitada. E essa redução no número dos respondentes pode ser um indicativo de eventual falta de acessibilidade.

Sem contemplar aqui todas as razões para esse comportamento dos respondentes, trago algumas afirmações de Maryanne Wolf, neurocientista pesquisadora, ao refletir sobre como o mundo digital está influenciando nos processos de leitura:

Os jovens cérebros leitores estão mudando sem que a maioria das pessoas se incomode, muito embora mais e mais dos nossos jovens leiam apenas aquilo que lhes é exigido, e muitas vezes nem mesmo isso: “MC; NL” (muito comprido; não li). [...] Num meio que nos defronta continuamente com um excesso de informações, a grande tentação de muitos é se retirar para depósitos conhecidos de informações facilmente digeríveis, menos densas, intelectualmente menos exigentes. A ilusão de estarmos informados por um dilúvio diário de informações dimensionadas eletronicamente para o olho pode dificultar uma análise crítica de nossas realidades complexas. (WOLF, 2018, pp. 11 e 20)

Em trabalhos que têm por núcleo temático a ATT, cremos que essa realidade da leitura no mundo digital deve também ser considerada e avaliada. E isso já está sendo feito no Brasil, com o Projeto PorSimples⁸⁹, por meio das ferramentas SIMPLIFICA e FACILITA. Esta última simplifica conteúdos de *sites* da *web* automaticamente.

Por fim, cremos que o questionário, ainda que com um pequeno número de respondentes, validou os resultados do Estudo-Piloto 2, em que propusemos várias alternativas de simplificação a alguns trechos de Sentenças dos JECs. As reescritas consideradas mais simples nesse Estudo-Piloto 2 foram também assim consideradas pelos respondentes do questionário. Validou também os resultados advindos das análises feitas por ferramentas computacionais utilizadas em nossa pesquisa, nos Estudos-Pilotos 1.

⁸⁹ No Capítulo 5, comentamos a respeito dessas ferramentas.

No próximo capítulo, apresentamos mais dois estudos-pilotos realizados durante nossa pesquisa. Desta vez, nosso olhar se volta para a sintaxe presente em algumas Sentenças do CE, mais precisamente para o comportamento dos constituintes frasais relacionados ao verbo.

7 ESTUDOS PARCIAIS 2

Dando continuidade à apresentação dos estudos-pilotos que realizamos durante a pesquisa, neste capítulo apresentamos mais dois trabalhos. Desta vez, nosso olhar se volta para o comportamento sintático dos textos que conformam o nosso CE.

Iniciamos, então, pelo Estudo-Piloto 3, que analisou uma amostra de 110 Sentenças do CE.

7.1 ESTUDO-PILOTO 3 – CONSTITUINTES FRASAS DA AMOSTRA 1 (110 SENTENÇAS DO CE)

A fim de descrevermos mais pormenorizadamente o comportamento sintático dos textos que conformam o nosso CE, criamos uma amostra com 110 sentenças que dele fazem parte. Afinal, a sintaxe verificada nas frases das Sentenças nos parecia bastante complexa, merecendo uma abordagem à parte. No item 7.1.1, explicamos como procedemos para a compilação dessa amostra para um estudo específico. E, no item 7.1.2, comentaremos os dados das análises realizadas.

Essa verificação, consubstanciada em um Estudo-Piloto, visa iluminar uma outra dimensão do nosso objeto de estudo, o texto especializado do âmbito jurídico, o plano da sua organização sintática frasal e textual. Nosso objetivo, nesse estudo parcial, é antever o quanto e como essa organização sintática tenderia a influenciar na tessitura da complexidade e a repercutir na caracterização de um *modus dicendi* peculiar.

7.1.1 Constituição da amostra do CE

Para a compilação de uma amostra do nosso CE, doravante Amostra 1, selecionamos um conjunto de 110 Sentenças que contemplassem todos os critérios utilizados na compilação da totalidade do CE: natureza da demanda, período de coleta e Turma Recursal. A seguir, apresentamos três tabelas: a Tabela 7.1 traz a distribuição das Sentenças por natureza da demanda, a Tabela 7.2 apresenta a distribuição da amostra por Turma Recursal, e a Tabela 7.3 mostra a distribuição da amostra por período de coleta.

Na última linha da Tabela 7.1, nossa amostra de 110 Sentenças representa um percentual de 25% do nosso CE. E, para cada temática, selecionamos um número de Sentenças cujo percentual ficasse próximo desse percentual de 25%. A única exceção foi

a temática Acidente de Trânsito, que, com 6 Sentenças, representou 40% da totalidade de sentenças relativa a esse tema.

Tabela 7.1 – Total de Sentenças por Temática e Quantidade de Sentenças para a Amostra 1

TEMÁTICA DA SENTENÇA	TOTAL DE SENTENÇAS DO CE	TOTAL DE SENTENÇAS DA AMOSTRA 1	PERCENTUAL DA AMOSTRA 1 EM RELAÇÃO AO CE
Água	45	12	26,66%
Acidente de Trânsito	15	6	40,00%
Energia Elétrica	73	17	23,29%
Inadimplência	55	13	23,63%
Plano de Saúde	68	17	25,00%
Prestação de Serviços	99	25	25,55%
Telefonia	85	20	23,53%
TOTAL	440	110	25,00%

A Tabela 7.2 apresenta a distribuição da amostra por Turma Recursal.

Tabela 7.2 – Total de Sentenças por Turma Recursal

TURMA RECURSAL CÍVEL	TOTAL DE SENTENÇAS DO CE	TOTAL DE SENTENÇAS DA AMOSTRA 1	PERCENTUAL DA AMOSTRA 1 EM RELAÇÃO AO CE
1TRC	110	29	26,36%
2TRC	110	30	27,27%
3TRC	110	24	21,81%
4TRC	110	27	24,54%
TOTAL	440	110	25,00%

A Tabela 7.3 apresenta a distribuição da amostra por período.

Tabela 7.3 – Total de Sentenças por Ano de Coleta

ANO DE COLETA	TOTAL DE SENTENÇAS DO CE	TOTAL DE SENTENÇAS DA AMOSTRA 1	PERCENTUAL DA AMOSTRA 1 EM RELAÇÃO AO CE
2018	220	47	21,37%
2019	220	63	28,63%
TOTAL	440	110	25,00%

Reunindo as sentenças da Amostra 1 num único arquivo, geramos um documento com 141.047 palavras, segundo a contagem do editor de texto Word. Em relação às 547.602 palavras do CE, o total de palavras da amostra corresponde a um percentual médio de 25%. A Tabela 7.4 traz mais alguns dados do CE e da Amostra 1 fixada.

Tabela 7.4 – Dados do CE e da Amostra 1

	CORPUS DE ESTUDO (CE)	AMOSTRA 1
TOTAL DE TEXTOS	440	110
TOTAL DE FRASES	31.601,52	7.491
PALAVRAS POR FRASE	18,22	19,25
TOTAL DE <i>TOKENS</i>	543.407	137.700
TOTAL DE <i>TYPES</i>	15.388	8.477
HAPAX	5.258	3.239

A ferramenta utilizada para a geração do número de *tokens* (palavras do texto), de *types* (palavras diferentes do texto) e de *hapax legomenon* (palavras de frequência única no texto) foi o sistema AntConc 3.2.1w⁹⁰. Para verificar o total de frases e o total de palavras por frase, utilizamos o sistema NILC-Metrix⁹¹ (Coh-Metrix 3.0, nova versão de 2020).

A Tabela 7.4 permite verificar que, em relação ao CE, a Amostra 1 apresenta percentuais próximos a 25% no que se refere a número de: textos (25%), *tokens* (24,3%) e frases (23,7%). O número de palavras por frase da amostra, por sua vez, é semelhante

⁹⁰ Sistema AntConc 3.2.1w. Disponível em <http://www.laurenceanthony.net/software.html>

⁹¹ Sistema NILC-Metrix. Disponível em: <http://fw.nilc.icmc.usp.br:23380/nilcmetrix>

ao do CE, com apenas 1 palavra a mais. Em relação ao *hapax legomenon*, que evidencia as especificidades e as preferências lexicais do autor de um texto, o percentual de ocorrência ficou muito semelhante nos dois conjuntos de textos: 34,17% no CE (5.258 para 15.385 *types*); e 38,20% na amostra (3.239 para 8.477 *types*). Tais dados permitem concluir pela representatividade da Amostra 1 em relação ao CE nesses quesitos. Salientamos, porém, que esses dados não influíram na fixação prévia da amostra.

Cumprе referir que a representatividade do tamanho de um *corpus* é tema sempre controverso, foco de acalorados debates. Na seção 4.2, já tratamos desse tema, mas o retomamos aqui com a seguinte afirmação de Biber (2012, p. 12):

Os livros sobre teorias de amostras, no entanto, ressaltam que o tamanho da amostra não é o fator mais importante na seleção de uma amostra representativa [...] A representatividade se refere ao quanto uma amostra inclui de toda a gama de variabilidade de uma população.

Entendemos por população da Amostra 1 o nosso CE. Dessa forma, estando ela composta por proporções mais ou menos semelhantes de cada temática do CE, e conforme os dados apresentados na Tabela 7.4, acreditamos que pode ser considerada efetivamente representativa dessa “população de textos”.

Mesmo não sendo o caso da Amostra 1, cabe ainda destacar, quanto à questão do tamanho de um *corpus*, que existem ferramentas criadas no âmbito dos estudos relativos ao Processamento da Linguagem Natural (PLN) que permitem comparar *corpora*, mesmo que sejam de diferentes tamanhos. É o caso da ferramenta EXATOLp (LOPES, 2012), utilizada por FINATTO (2018, p. 358). Esta pesquisadora utilizou uma métrica do EXATOLp para comparar uma amostra de *blogs* sobre Pneumopatias Ocupacionais (764 frases e 16.591 palavras) com uma amostra de artigos acadêmicos sobre a mesma temática (9.239 frases e 241.806 palavras). O índice tf-dcf (índice comparativo utilizado) do EXATOLp não considera o número de ocorrências brutas dos termos no *corpus* contrastante, mas, sim, o logaritmo deste número de ocorrências. Assim, com essas ponderações matemáticas, segundo Finatto (2017, p. 367), é possível

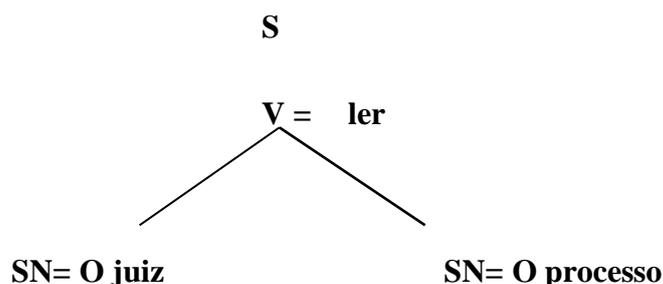
validar a ideia de que, conforme o dito popular, “tamanho não é documento”, e isso também vale para *corpus* e *corpora*. O fundamento da LC sobre o *corpus* extenso ser representativo segue válido, mas também se podem obter ótimos indicativos sobre padrões lexicais e terminológicos com amostras pequenas.

Na próxima seção, confiando que a Amostra 1 é suficientemente representativa para o fim a que nos propomos - a descrição da sintaxe presente nos textos -, explicamos os procedimentos para as análises de nossa amostra de 110 textos de Sentenças.

7.1.2 Processamento e Análise dos Dados

Inicialmente, analisamos os constituintes das sentenças⁹² de nossa amostra de Sentenças, porque eles são uma das variáveis analisadas em estudos sobre a complexidade textual e, mais especificamente sentencial – frasal e oracional. Entendemos como constituintes da sentença os elementos da oração que formam com o verbo uma estrutura de argumentos ou valência verbal. Para exemplificar este entendimento, apresentamos, tal como Zilio (2015, p. 32), um exemplo de sentença estruturada a partir do verbo (V), ou sintagma verbal (SV), partindo para os seus constituintes, os sintagmas nominais (SNs):

(7.1) **O juiz leu o processo.**



Este autor explica que essa representação difere da representação mais tradicional na Linguística, que separa sujeito e predicado com base nos conceitos de tema-rema. Justifica a opção pela representação centrada no verbo por ser mais prática e necessária a um tratamento computacional. Um tratamento computacional como esse, focado na sintaxe do texto, permite lidar com uma grande quantidade de dados (caso dos nossos *corpora*) de modo mais rápido.

Cumprе referir, entretanto, que as análises que realizamos foram manuais. Ou seja, lemos cada uma das orações presentes em nossa amostra e delas extraímos os constituintes frasais com função de sujeito e de complementos verbais. Inicialmente pensamos em utilizar o *parser* PALAVRAS (BICK, 2000), um etiquetador

⁹² Sentença aqui é entendida como uma sequência de palavras iniciada com uma letra maiúscula e encerrada com uma pontuação final. E uma oração pode ser uma sentença ou parte dessa sentença que se organiza em torno de um verbo ou locução verbal.

morfossintático e gramatical que “adota o verbo como cabeça de sentença e é capaz de gerar representações em forma de árvores” (ZILIO, 2015, p. 34).

O *parser* PALAVRAS, embora relativamente antigo, tem sido considerado um dos melhores sistemas de seu gênero (PARDO; NUNES, 2008) para o tratamento computacional da Língua Portuguesa, especialmente no âmbito do reconhecimento sintático. Em diferentes trabalhos, ao longo dos anos, tem-se reconhecido que ele “atinge um percentual de acerto de 97% no que concerne à árvore de dependências (organização hierárquica das funções sintáticas)” (ZILIO, 2015, p. 34). Em Bick (2000), é possível encontrar uma descrição completa do funcionamento desse *parser*.

Ferramentas computacionais como o PALAVRAS analisam automaticamente elementos presentes em palavras, sentenças, textos e também em conjuntos de textos. Algumas dessas ferramentas apenas etiquetam determinados elementos como substantivos, adjetivos, verbos etc., e/ou fazem algum tipo de decomposição morfológica, como a lematização das palavras que perfazem um texto ou frase. Outras apresentam a estrutura hierárquica de uma sentença, as relações de dependência sintática entre os elementos de uma sentença, como é o caso do PALAVRAS.

Conforme mencionamos anteriormente, inicialmente pretendíamos utilizar o *parser* PALAVRAS (BICK, 2000) para as análises dos constituintes frasais de nossos textos. Porém, verificamos que essa ferramenta, apesar de ser muito útil e eficaz, apresentou alguns problemas nas análises processadas em nossa amostra. É o que se vê no seguinte exemplo. A frase que nos interessa é a que está em negrito, no Quadro 7.1, porém colocamos o trecho anterior, para melhor compreensão de todo o contexto.

Quadro 7.1 – Problemas de Análise no Parser Palavras

TRECHO DA SENTENÇA
FULANA DE TAL ingressa com a presente ação contra CORSAN alegando que é usuária dos serviços prestados pela demandada, sendo que na fatura com vencimento em janeiro foi lançado um valor de consumo que considera exagerado (R\$ 1.154,65). Tentou resolver o impasse diretamente com a demandada e não obteve êxito. Pagou a fatura com atraso, o que acarretou cobrança de juros e multa. Em emenda da inicial, refere que ficou sem abastecimento de água entre os dias 08 e 23/11/2018. Requer seja a demandada condenada a indenizar pelos danos morais experimentados, em valor de R\$ 10.200,00, além dos pedidos de revisão da fatura referida, com a exclusão dos valores excessivos, tomando-se por base os consumos anteriores. Tudo com a inversão do ônus da prova.
ANÁLISE DO PARSE PALAVRAS
Requer [Requer] <hum> <*> PROP M/F S @SUBJ > #1->2 (@SUBJ = sujeito)

seja	[ser] <fmc> <aux> V PR 3S SUBJ VFIN @FS-STA #2->0
a	[o] <artd> DET F S @>N #3->4
demandada	[demandado] ADJ F S @< PRED > #4->0
condenada	[condenar] <vH> <n> <mv> V PCP F S @ICL-AUX< #5->2
a	[a] PRP @<PIV #6->5
indenizar	[indenizar] <vH> <mv> V INF @ICL-P< #7->6
[...]	
tomando-	[tomar] <clb> <hyfen> <vH> <mv> V GER @ICL-<ADVL #38->7
se	[se] PERS M/F 3S/P ACC @<SUBJ #39->38
por	[por] PRP @<PIV #40->38
base	[base] <Labs> <geom> <part-build> <ac-cat> <cm-liq> <inst> N F S @P< #41->40
os	[o] <artd> DET M P @>N #42->43
consumos	[consumo] <act-d> N M P @< ACC > #43->38
anteriores	[anterior] <nh> <np-close> ADJ M P @N< #44->43
\$. #45->0	
</β>	

Vejamos três situações que nos pareceram incorretas na análise do Parser Palavras em relação ao trecho em negrito do Quadro 7.1.

a) Verbo **requer**:

Este verbo está numa forma nominal - infinitivo não flexionado -, mas tem um referente no trecho anterior: FULANA DE TAL. Pelo trecho se depreende que é a FULANA DE TAL que requer. Trata-se, portanto, de um caso de sujeito elíptico.

No Parser Palavras, este verbo foi considerado como um nome e a ele foi atribuída a função de sujeito (PROP M/F S @<**SUBJ**>), o que não é o caso.

b) Particípio **demandada**:

Este particípio, no trecho em questão, refere-se à entidade contra quem foi proposta a ação, a demanda, como se vê no trecho do Quadro 7.1. O termo **demandada** é utilizado para designar uma ação na Justiça. Nesse caso, seus atores são chamados de **demandante** e **demandado**. O primeiro é aquele que propõe a demanda; o segundo é aquele contra quem a demanda é proposta.

Então, nesta frase, o particípio **demandada** designa um dos atores da sentença, no caso, a CORSAN. Trata-se de um substantivo que atua como sujeito de **seja condenada a indenizar**.

O Parser Palavras analisou esta palavra como um adjetivo atuando como predicativo (ADJ F S @<PRED), o que também não é o caso.

c) A palavra **consumos**:

O substantivo **consumos** no trecho em questão atua como núcleo do sujeito do verbo **tomando-se**. O Parser Palavras analisou este substantivo como objeto direto (@<ACC).

Diante disso, como nossa análise exigia maior acuidade na observação e descrição dos dados, optamos por continuar selecionando os nossos constituintes frasais de forma manual.

Nossas análises em relação à amostra se deram sobre os constituintes das sentenças, porque, entre outros motivos, a extensão dos elementos constituintes das sentenças é uma das variáveis consideradas na avaliação da complexidade textual e, mais especificamente, em trabalhos que estudam a complexidade sentencial. A complexidade de uma sentença está relacionada a fatores como: extensão da sentença, extensão dos seus elementos constituintes, ordem dos elementos da sentença, presença de anáforas etc.

A seguir, no Quadro 7.2, apresentamos um trecho extraído de nossa amostra com análise de alguns fatores de potencial complexidade para um leitor adulto de escolaridade limitada.

Quadro 7.2 – Exemplo de frase complexa

No que pertine à preliminar de complexidade da causa, levantada pela empresa requerida, em sua contestação, (1) merece ser desde logo afastada (2), porquanto se mostra absolutamente desnecessária a realização de perícia técnica (3) quando os elementos de prova constantes dos autos são suficientes para o deslinde da controvérsia.

- (1) A frase começa numa ordem invertida, com um adjunto adverbial longo (parte sublinhada).
- (2) O sujeito de “merece ser desde logo afastada” está elíptico e expresso no adjunto adverbial que inicia o trecho (preliminar de complexidade da causa).
- (3) Esta oração também está numa ordem invertida: o sujeito (a realização de perícia técnica) está depois do verbo (se mostra). E a conjunção “porquanto”, que inicia esta oração, não é muito usual.
- (4) O Índice Flesch do trecho ficou negativo -7,51. Ou seja, é um texto extremamente difícil.

Na Tabela 7.5, apresentamos algumas reescritas simplificadas para o parágrafo, com a medição do Índice Flesch⁹³ ao lado.

⁹³ Como já observamos neste trabalho, o Índice Flesch contempla apenas a extensão das palavras e das frases e por isso não pode ser tomado isoladamente. De todo modo, ele ajuda no processo inicial de verificação da complexidade de um texto.

Tabela 7.5 – Reescritas Simplificadas da Frase Constante no Quadro 7.2

Reescritas	IF
Inicialmente, afasto a preliminar de complexidade da causa trazida pela empresa requerida, pois há elementos de prova suficientes para resolver o caso.	7,31
Inicialmente, afasto a preliminar de complexidade da causa, pois há elementos de prova suficientes para resolver o caso.	19,06
Inicialmente, afasto a preliminar de complexidade da causa trazida pela empresa requerida, pois os elementos de prova que constam nos autos são suficientes para resolver o caso.	20,89
Inicialmente, afasto a preliminar de complexidade da causa trazida pela empresa requerida. Os elementos de prova que constam nos autos são suficientes para resolver o caso.	30,65
Inicialmente, afasto a preliminar de complexidade da causa, pois há elementos de prova suficientes para resolver o caso. A preliminar de complexidade da causa tem a ver com a necessidade de se fazer uma perícia técnica para provar o que se discute no processo.	43,84

Colocamos as reescritas em ordem crescente de facilidade na leitura, conforme estimada pelo Índice Flesch. Podemos observar que as reescritas que dividiram a sentença em mais de uma foram as que mais facilitaram o trecho, confirmando o que vários estudos anteriores sobre a complexidade textual já evidenciaram. É o caso, por exemplo, do Projeto PorSimples, em que, segundo Leal (2019), os processos mais utilizados na simplificação textual foram a divisão de sentenças e a substituição de palavras.

Este pesquisador afirma que a análise e avaliação de cada sentença de um texto permite uma informação mais exata dos seus pontos complexos, o que tenderia a auxiliar um redator interessado na tarefa de simplificação textual. Ainda segundo este autor, “os primeiros avanços na área de simplificação textual automática surgiram com a ideia de dividir sentenças longas em sentenças menores para melhorar os resultados dos analisadores sintáticos” (LEAL, 2019, p. 25).

No trabalho de Leal (2019), são relatados uma série de estudos que se debruçaram sobre a questão da complexidade sentencial. Entre os trabalhos citados, os de Dell’Orletta *et al.* (2011, 2014) concluíram que a avaliação do nível sentencial é bem mais complexa que a avaliação global de um documento. Verificaram ainda que o tamanho das sentenças e o tamanho das palavras foram os dois primeiros indicadores mais avaliados para medição da complexidade.

Entre as características sintáticas analisadas nesses trabalhos citados por Leal (2019), destacamos as que seguem:

- **processos de coordenação** – quantidade de orações coordenadas e de conjunções coordenativas;

- **processos de subordinação** – distribuição de orações subordinadas *versus* principais, extensão das cadeias subordinadas, ordem da oração subordinada em relação à principal;

- **profundidade da árvore sintática** – extensão desde a raiz até a última folha, extensão das cadeias de complementos preposicionais ou modificadores nominais e adjetivos;

- **dependências sintáticas** - número de palavras entre a raiz sintática e o dependente, o total de ligações;

- **ordem dos constituintes na oração** - ordenação relativa do sujeito e objeto em relação ao verbo principal.

Creemos que a nossa análise manual, um a um, dos argumentos verbais permitirá uma melhor visibilidade da profundidade das árvores sintática e das dependências sintáticas dos textos de nossa amostra. Tais fatores, como vimos acima, são considerados na aferição da complexidade sentencial. Portanto, acreditamos que os dados de nossas análises podem nos ajudar a traçar um perfil da complexidade sintática envolvida nas sentenças do nosso CE. Com isso, pode-se subsidiar, com dados mais precisos, uma tarefa posterior de simplificação textual.

Na próxima seção apresentamos os dados sobre os constituintes frasais com função de sujeito, doravante CFSs, presentes em nossa amostra, conforme os verificamos.

7.1.2.1 Constituintes Frasais com Função de Sujeito (CFSs)

Antes de apresentarmos os dados relativos aos CFSs presentes em nossa amostra, retomamos aqui que nossas análises em relação aos constituintes da oração partem do verbo, dos elementos que a ele estão relacionados. Assim, nossa análise do sujeito aqui não será a que se dá entre sujeito e predicado, ou entre tema e rema, como já afirmamos.

Nossa atenção está voltada em como se organizam as orações presentes em nossos textos em relação aos constituintes dos verbos: de quantos elementos se constituem esses sintagmas com função de sujeito, em que posição se encontram em relação ao verbo, antes

ou depois, se estão expressos na oração ou precisam ser depreendidos mediante um processo de anáfora, etc.

De todo modo, considerando que os elementos que se relacionam ao verbo se diferem em termos de função, julgamos importante esclarecer aqui o que entendemos em relação à função de sujeito.

Entendemos, tal como Perini (1994, p. 72), que o sujeito é o constituinte da oração que está em relação de concordância com o núcleo do predicado, que é o verbo. Este autor esclarece que entender o sujeito como o elemento que entretém uma relação de concordância com o verbo não significa uma negação de outras noções relacionadas ao sujeito, tal como a de agente, paciente, tópico. Perini explica que tais propriedades devem ser analisadas

dentro de componentes separados da gramática: a noção de sujeito se formula em termos de relações formais (distribucionais, relacionais, etc.); a de agente em termos do conteúdo nocional; e a de tópico em termos de inserção em um contexto de comunicação. (PERINI, 1994, p. 74)

Assim, nosso trabalho, sem negar outras formas de análise do elemento que tem a função de sujeito na oração, centra-se na identificação e descrição de como se apresenta esse constituinte na oração. Também retomamos que a nossa análise dos CFSs, como já dissemos, foi manual. Ou seja, lemos cada uma das orações presentes nos textos de nossa amostra e fomos extraíndo os CFSs que exerciam a função de sujeito, classificando-os, inicialmente, em dois grandes grupos: sujeitos realizados e sujeitos elípticos. Os realizados foram classificados em: pré-verbal (sujeitos colocados antes do verbo); pós-verbal (sujeitos colocados após o verbo); pronominal (sujeitos representados por pronomes – relativos, demonstrativos, retos, etc.); e oracional (oração como sujeito de outro verbo). Os elípticos foram classificados em duas situações – quando determinado pela flexão número-pessoa do verbo ou por sua presença em alguma oração antecedente.

A classificação quanto à posição do sujeito – pré ou pós-verbal – deu-se apenas em relação aos CFSs constituídos por sintagmas nominais.

Os sujeitos pré-verbais encaixam-se no padrão da ordem canônica natural da Língua Portuguesa – sujeito-verbo-objeto. Por isso, em princípio, as orações das quais eles fazem parte podem ser mais fáceis. E os sujeitos pós-verbais, por quebrarem essa ordem canônica da língua, tornariam, em princípio, as orações de que fazem parte, mais difíceis. A ordem desses constituintes é, como vimos, variável considerada nos estudos de aferição da complexidade sentencial.

A hipótese quanto à expectativa do leitor em relação à organização sintática da frase é uma das utilizadas em estudos sobre o letramento para sustentar que sentenças que fogem ao padrão canônico podem ser um fator de complexidade leitora.

Bailin e Grafstein (2016) afirmam que a conexão das unidades de informação de um texto é uma das variáveis para se analisar a inteligibilidade desse texto. Segundo esses autores, quando a ordem natural da gramática da língua é alterada, a compreensão leitora pode ser afetada. O leitor, no ato da leitura, busca, em seus conhecimentos linguísticos, estruturas sintagmáticas que lhe sejam mais conhecidas e, em encontrando, este conhecimento prévio favorece a compreensão. É o que diz Coscarelli (2002, p. 14)

Da mesma forma que existe um padrão silábico mais frequentemente encontrado em cada língua, existem também as estruturas sintáticas mais usadas. A estrutura mais simples do português é aquela em que se tem o sujeito seguido do verbo que, por sua vez, é acompanhado por um complemento.

Kleiman (2002) também coloca a ruptura da ordem canônica mais usual da língua como um fator dificultador no processamento da leitura. Porém, observa que, por vezes, colocar no início da frase uma informação conhecida do leitor, mesmo numa estrutura considerada mais complexa, pode facilitar mais a compreensão que a ordem canônica. É o que pode ocorrer, por exemplo, numa estrutura passiva que inicia com um elemento dado, já conhecido do leitor, dando mais fluidez à progressão temática do texto. Ela traz o seguinte exemplo:

A Terra está mostrando as terríveis marcas das feridas que lhe fizemos, e está lançando seu S.O.S. E preciso ouvir esse apelo, e acudir com urgência. Se este apelo não for ouvido, estaremos condenando nossa biosfera à degeneração, com os ambientes aéreo, terrestre e aquático comprometidos. (KLEIMAN, 2002, p. 40)

Na parte sublinhada do exemplo tal, acima, Kleiman afirma que “este apelo”, objeto na frase anterior, como sujeito numa construção passiva, torna o texto mais fácil porque a frase inicia com uma informação já dada no texto e, portanto, já conhecida do leitor. De todo modo, em nossas análises, “este apelo” seria classificado como pré-verbal, porque vem antes do verbo.

Quanto aos sujeitos pronominais, cabe lembrar que também são tidos, nas pesquisas sobre leitura e letramento, como dificultadores da leitura compreensiva, porque as referências pronominais podem se referir a mais de um elemento expresso no texto. Nessas situações, o leitor é forçado a buscar outras pistas no texto para estabelecer a relação anafórica corretamente, atrasando a leitura e a fluência na compreensão do que é

lido. É o que ocorre com o pronome “ele”, no exemplo (7.2), a seguir, extraído de nossa amostra.

- (7.2) A testemunha Fulano de Tal, em seu depoimento, declarou que o autor procurou o local onde **ele** trabalhava para fazer passaporte, visto, bem como auxílio na compra do bilhete de viagem. **Afirmou** ao requerente que não tinha experiência **naquilo**, que nunca havia comprado passagem de avião.

O pronome “ele” no exemplo (7.2) pode se referir à “testemunha Fulano de Tal” como também a “o autor”. O leitor terá de procurar as pistas contextuais para compreender que este “ele” só pode ser a “testemunha Fulano de Tal”, pois, em princípio, “o autor” não iria até o seu local de trabalho apenas para fazer o passaporte. Sublinhamos também o pronome “naquilo”, que igualmente não deixa muito claro o seu referente.

Ainda neste exemplo, destacamos em negrito a forma verbal “afirmou”, que retoma, por um processo de elipse, “A testemunha Fulano de Tal”, sujeito da primeira oração da frase anterior. Trata-se de um caso de sujeito elíptico, outra classificação dos sujeitos da nossa amostra. Nesse sentido, observamos que, apesar de nossa análise centrar-se no constituinte expreso na oração, julgamos pertinente relacionar os casos de sujeitos elípticos porque, mesmo não estando explicitamente colocados na oração, podem ser retomados por um processo de referenciação situacional e textual. E esse processo, para um leitor não muito proficiente, pode ser um fator de complexidade.

Cumprе referir, quanto ao sujeito elíptico ou oculto, como também é chamado, que Perini (1994) questiona a atribuição de traços de pessoa e número a um elemento vazio, como em casos como o exemplo (7.3), extraído de nossa amostra.

- (7.3) Passo a fundamentar.

Para este autor, o exemplo (7.3) não tem sujeito, porque “as formas pessoais do verbo são geradas livremente e [...] é aceitável não porque haja aí concordância, mas antes porque não sofre a ação de nenhum filtro” (PERINI, 1994, p. 90).

Apesar disso, entendemos que, no contexto dos nossos textos, é possível inferir, pela situação comunicativa, o elemento que está em relação de concordância com a locução “passo a fundamentar”, que seria o julgador que está apreciando o caso em questão. Em nossa amostra, a grande maioria desse tipo de sujeito elíptico é representada por verbos que retomam o redator da sentença, como os verbos em negrito nos exemplos a seguir:

- (7.4) **Tenho adotado** o entendimento de que cabe à empresa prestadora do serviço provar o ocorrido.
- (7.5) ISTO POSTO, **julgo** PARCIALMENTE PROCEDENTE os pedidos [...].
- (7.6) Assim, **penso** que uma indenização no valor pleiteado.

E, por último, identificamos a presença de sujeitos oracionais, que representam, segundo Perini (1994, p. 204), sintagmas nominais complexos. A grande maioria deles aparece, em nossa amostra, após o verbo, como no exemplo (7.7), a seguir:

- (7.7) É importante [**esclarecer**] [que comprador e consumidor são termos juridicamente distintos, assim como vendedor e fornecedor].

O infinitivo “esclarecer” atua como sujeito da forma verbal “é”, da primeira oração e é complementado, formando um SN complexo, pela oração entre colchetes.

Entre os tipos de sujeitos apontados pela gramática, existem os chamados “sujeitos indeterminados”, que não foram por nós considerados nesta análise. Diferentemente dos sujeitos elípticos, que não estão expressos na oração, mas podem ser depreendidos no texto ou na situação comunicativa, os sujeitos indeterminados estão relacionados a um verbo, mas não é possível detectá-lo no texto ou na situação comunicativa. Sua interpretação tem muito mais um caráter semântico que sintático. Por isso, neste estudo, não foram objeto de análise.

Tendo esclarecido o nosso ponto de vista e a nossa porta de entrada para as análises dos constituintes oracionais com função de sujeito em nossa amostra, passamos a apresentar dados quantitativos de cada um dos tipos selecionados.

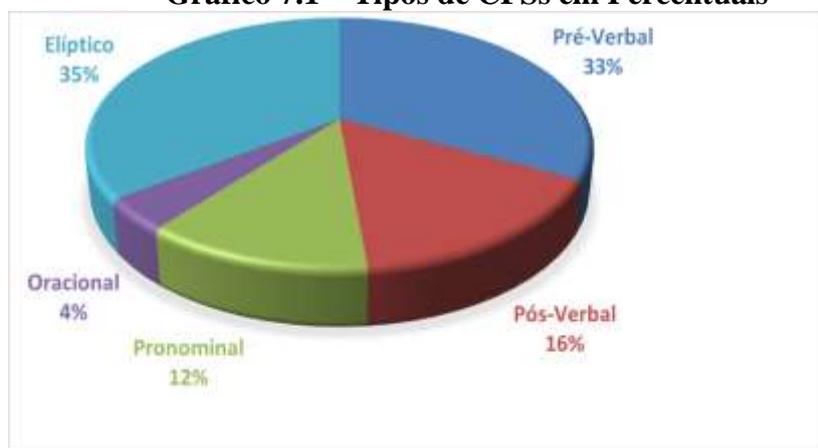
Na Tabela 7.6, apresentamos o total de CFSs com a quantidade encontrada para cada tipo.

Tabela 7.6 – Ocorrências por Tipos de CFCVs da Amostra

TIPOS DE SUJEITOS					
	ELÍPTICOS	REALIZADOS			
		SINTAGMA NOMINAL		PRONOMINAL	ORACIONAL
		Pré-Verbal	3.869		
		Pós-Verbal	1.884		
			5.753	1.459	486
SUBTOTAL	4.159				7.698
TOTAL					11.857

Com os dados da Tabela 7.6, geramos, no Excel, o Gráfico 7.1, com o percentual respectivo a cada tipo de sujeito.

Gráfico 7.1 – Tipos de CFSs em Percentuais



Conforme os dados da Tabela 7.6 e do Gráfico 7.1, verificamos que os sujeitos elípticos são os de maior ocorrência na amostra composta por 110 textos, seguidos dos sujeitos pré-verbais e pós-verbais. Tais dados permitem-nos, de pronto, fazer algumas inferências aqui.

Como já observamos, sujeitos elípticos podem ser considerados fator de maior complexidade leitora porque requerem uma sobrecarga maior de memória de trabalho. Assim sendo, poderíamos dizer que os textos da nossa amostra, com sujeitos elípticos em maior percentual, demandam um trabalho a mais na leitura e, por isso, são complexos ao nosso perfil de leitor.

Se juntarmos o percentual de sujeitos elípticos (35%) com o percentual de sujeitos pós-verbais (16%), teremos um percentual de 51%. Como também já mencionamos, os sujeitos pós-verbais, por não estarem na ordem canônica da língua, podem ser considerados fator de maior complexidade para a leitura. Ou seja, metade dos sujeitos de nossa amostra apresentam características que aumentam a complexidade textual. Essa complexidade aumenta ainda mais se acrescentarmos, a esse percentual de 51% relativo aos sujeitos elípticos e pós-verbais, o percentual de sujeitos pronominais. Teremos, com essa soma, um percentual de 63% de sujeitos que demandam bastante trabalho do leitor.

Após essa análise inicial, adotamos dois procedimentos.

Procedimento 1:

Distribuímos os sujeitos pré-verbais, pós-verbais e oracionais por número de palavras. Assim procedemos, para verificar a extensão das dependências sintáticas, ou seja, o número de palavras entre a raiz sintática (no caso, o SV) e o dependente (no caso, o SN com função de sujeito).

Para Fulgêncio e Liberato (2010, p. 139), o mais importante não é a medida do comprimento da sentença, mas dos constituintes da sentença individualmente, pois são eles que formam unidades de significado. Além disso, como já observamos, afora a extensão das dependências sintáticas – uma das variáveis consideradas nos estudos relativos à complexidade sentencial –, a quantidade de palavras de uma sentença como um todo é variável considerada em muitas das métricas utilizadas para aferição da inteligibilidade textual.

A seguir, apresentamos esses dados por tipos de sujeito.

a) Sujeitos Pré-Verbais

A Tabela 7.7 apresenta as ocorrências dos sujeitos pré-verbais por número de palavras.

Tabela 7.7 – CFSs Pré-Verbais por Número de Palavras

QUANTIDADE PALAVRAS	OCORRÊNCIAS	PERCENTUAL DE OCORRÊNCIAS
1 a 10	3654	94%
11 a 20	176	5%
21 a 30	31	0,80%
30 ou +	8	0,20%
TOTAL	3869	100%

Pela Tabela 7.7, podemos verificar que 94% dos sujeitos pré-verbais são constituídos de no máximo 10 palavras. Porém, muitos deles apresentam seus núcleos expandidos por encaixes oracionais, como se vê nos exemplos a seguir.

- (7.8) **A manutenção [de que trata este artigo]** é extensiva, obrigatoriamente, a todo o grupo familiar inscrito quando da vigência do contrato de trabalho.
- (7.9) **A questão [pela qual se insurge a autora]**, (sic) tem relação com a necessidade de uso dos dados pessoais de seu cartão de crédito, sendo incontroverso que a autora desistiu da prestação dos serviços logo após a contratação.

- (7.10) **Digressões abstratas [de que o procedimento esteja excluído] não representa** (sic) prova efetiva que autoriza a falta de cobertura securitária.

O exemplo (7.8) contém 7 palavras; e os exemplos (7.9) e (7.10), 8 palavras. Apesar disso, em todos eles, a oração relativa encaixada sobrecarrega a memória de trabalho do leitor, porque ele precisa processar várias informações antes de chegar à principal. E essa sobrecarga na memória de trabalho pode prejudicar a compreensão. Encaixes como os exemplificados podem ser considerados inserções, que corresponderiam, segundo Liberato & Fulgêncio (2010, p. 130), a “uma ‘quebra’ na sentença, provocada por um trecho que é interposto numa estrutura linear e que interrompe a sequência esperada dos constituintes”. A estrutura linear esperada é a estrutura canônica da língua: sujeito-verbo-objeto. Perini (1994) observa que, na fala, essas quebras da estrutura linear são compensadas pela repetição dos elementos já mencionados, o que não ocorre na escrita.

O exemplo (7.10), com a falta de concordância verbal “não representa” ao invés de “não representam”, evidencia essa situação. A memória de curto prazo, ao ter de processar a informação trazida pela encaixada [**de que o procedimento esteja excluído**], não conseguiu manter o núcleo plural do sujeito – *digressões* –, e aparentemente relacionou o verbo com o sujeito da encaixada: procedimento.

Cumpram também referir que, em todas as orações relativas encaixadas dos exemplos (7.8), (7.9) e (7.10), o sujeito da principal transforma-se em objeto, sendo alterado o seu papel temático. Com isso, além de a representação do sujeito ter de ser mantida na memória de trabalho até chegar ao final da encaixada, a associação de um único item lexical a dois papéis temáticos diferentes, segundo Just e Carpenter (1992, p. 128), representa uma complexidade para a compreensão leitora.

Abaixo trazemos o exemplo (7.11), com mais de 30 palavras. São em menor número, mas julgamos pertinente fazer algumas observações.

- (7.11) De fato, embora não exista contrato expresso que demonstre o negócio jurídico realizado entre as partes, concluo que **os depósitos [efetuados pelo réu em benefício do autor (folha 34), [aliados às fotos de folhas 42 e 43 e aos e-mails trocados entre as partes (folhas 28, 30 e 31), em especial, o e-mail de folha 28] [no qual o requerido confirma] [que os pagamentos seriam realizados de acordo com a realização da Obra], corroboram** a versão inicial, no sentido de que autor e requerido ajustaram a construção de uma casa, que seria paga de modo parcelado até o final da obra.

No exemplo (7.11), entre o núcleo do sujeito – *depósito* – e o seu verbo – *corroboram* –, há 53 palavras, o que, por si só, evidencia o quanto o leitor terá de guardar na memória de trabalho até completar a ideia dessa oração. Se considerarmos os participípios “efetuados” e “aliados”⁹⁴ como orações restritivas reduzidas, essas 53 palavras estão organizadas em 4 orações que deverão ser processadas pelo leitor até que ele chegue ao verbo do sujeito inicial.

Para Santos (2013, p. 78),

inserções longas ou numerosas obrigam o leitor a proceder a mais fatiamentos, o que implica maior chance de ocorrer o esvaziamento da memória de trabalho antes mesmo de o leitor ter completado a construção do sentido local de todo o período.

b) CFSs Pós-Verbais

A Tabela 7.8 apresenta as ocorrências dos sujeitos pós-verbais por número de palavras.

Tabela 7.8 – CFSs Pós-Verbais por Número de Palavras

QUANTIDADE PALAVRAS	NÚMERO DE OCORRÊNCIAS	PERCENTUAL DE OCORRÊNCIAS
1 a 10	1453	77%
11 a 20	320	17%
21 a 30	77	4%
30 ou +	34	2%
TOTAL	1884	100%

Os sujeitos pós-verbais também apresentam, em sua maioria (77%), entre 1 e 10 palavras. Em segundo lugar, estão os sujeitos com 11 a 20 palavras, e 34 sujeitos estão constituídos de 30 ou mais palavras.

Abaixo trazemos alguns exemplos:

(7.12) O Tribunal *a quo* negou provimento ao apelo interposto pelo ora agravante, sob o fundamento de que, nas relações de consumo, as cláusulas limitativas de direito serão sempre interpretadas a favor do consumidor, desse modo, ao assim decidir, adotou posicionamento consentâneo com a jurisprudência desta egrégia Corte, que se orienta no sentido de considerar que, em se

⁹⁴ Segundo Perini (1994), essas construções não são orações, porque o comportamento desses participípios é idêntico ao de um adjetivo. Ainda que se possam considerar tais participípios como simples adjetivos, as estruturas em que eles se encontram são complexas, o que demandam maior trabalho na leitura, foco do nosso trabalho.

tratando de contrato de adesão submetido às regras do CDC, a interpretação de suas cláusulas deve ser feita da maneira mais favorável ao consumidor, bem como devem ser consideradas abusivas **as cláusulas [que visam a restringir procedimentos médicos]**.

- (7.13) Devido, pois, **o reembolso daquilo [que o contratante despendeu com o procedimento] [cuja cobertura foi negada pela seguradora]**.
- (7.14) Neste contexto, sem desconhecer o teor da Súmula 29 das Turmas Recursais, entendo comprovada **a falha na prestação dos serviços telefônicos a cargo da Empresa/Requerida bem como, a partir disso e da prova oral coligida (folha173), a configuração do abalo extrapatrimonial suportado pela Autora/consumidora [que, inegavelmente, teve sua rotina comercial alterada pela indisponibilidade do serviço telefônico, de cunho essencial para a atividade desenvolvida (ótica e relojoaria)] [refletindo, por via de consequência, negativamente no faturamento mensal] [que dela decorre]**.

O sujeito do exemplo (7.12) tem seu núcleo expandido por uma oração relativa encaixada. Entretanto, como o encaixe está à direita, isso demanda menos custo de processamento. Não há aqui uma quebra na sequência dos constituintes frasais, mas apenas uma extensão que delimita o sentido do núcleo. De todo modo, chama a atenção a extensão do período em que se encontra essa oração: 96 palavras organizadas em várias orações subordinadas encaixadas, o que, por si só, sobrecarrega a memória de trabalho na leitura, dificultando a compreensão. Fulgêncio e Liberato (2010, p. 138) afirmam:

Em geral, não é propriamente o comprimento excessivo de uma sentença o seu problema maior, e sim as características daí decorrentes. Na maioria das vezes, quando se tenta organizar muitas ideias numa só sentença – o que certamente lhe aumentará o tamanho –, não se tem como evitar inserções, anáforas ambíguas, estruturas passivas, enfim, características que dificultam o processamento.

No exemplo (7.13), o complemento do núcleo “daquilo” é uma oração relativa encaixada que, por sua vez, também apresenta uma palavra expandida por outra oração relativa encaixada, esta iniciada pelo pronome “cuja”, pouco usual no cotidiano da Língua Portuguesa. Ou seja, é um constituinte complexo com vários encaixes hierarquicamente diferentes e apresenta também um nexos pouco frequente na língua. Ainda segundo Fulgêncio e Liberato (2010, p. 141), “a quantidade de encaixamentos é outro ponto que causa a complexidade das estruturas e faz com que elas sejam mais difíceis de serem processadas”.

E o exemplo (7.14) apresenta outro exemplo de período longo: uma frase só com 79 palavras. O sujeito do período, em negrito, contém, só ele, 64 palavras. Trata-se de um

sujeito com dois núcleos (sublinhados no trecho), ambos ampliados por vários elementos - palavras e orações - em vários processos de subordinação com diferentes níveis de hierarquia. Em princípio, além de estar numa ordem diferente da canônica, estes sujeitos se apresentam numa configuração que tende a sobrecarregar a memória de trabalho e, conseqüentemente, dificultar a leitura.

c) CFSs Oracionais

Os sujeitos oracionais são representados por estruturas oracionais⁹⁵ que atuam como sujeito de outro verbo. A Tabela 7.9 apresenta a distribuição dos sujeitos oracionais da amostra por número de palavras.

Tabela 7.9 – CFSs Oracionais por Número de Palavras

QUANTIDADE PALAVRAS	NÚMERO DE OCORRÊNCIAS	PERCENTUAL DE OCORRÊNCIAS
1 a 10	151	31,06%
11 a 20	186	38,27%
21 a 30	82	16,87%
31 ou +	67	13,80%
TOTAL	486	100%

Os sujeitos oracionais de maior ocorrência estão na faixa de 11 a 20 palavras. Tal dado tem coerência, uma vez que os sujeitos são representados por orações. Após, em segundo lugar, estão os que contêm 1 a 10 palavras, seguidos dos que contêm 21 a 30 palavras. Na última faixa, de 31 ou mais palavras, ocorrem 67 casos.

A seguir apresentamos alguns exemplos desses sujeitos:

- (7.15) De outro lado, os planos de saúde apenas podem estabelecer para quais moléstias oferecerão cobertura, não lhes cabendo **limitar o tipo de tratamento que será prescrito**, incumbência essa que pertence ao profissional da medicina que assiste o paciente.
- (7.16) Inicialmente, não se desconhece **que está em processamento pedido de recuperação judicial da sociedade empresária impugnante, bem como que há recomendação, [ressalvado entendimento jurisprudencial diverso], no sentido de suspensão dos feitos executórios, [em que a referida empresa figure como demandada]**.
- (7.17) No caso concreto, verifica-se **que não se constata a verossimilhança das alegações da autora quanto à alegada falha por parte da empresa**

⁹⁵ Utilizamos a classificação da NGB na consideração dos sujeitos oracionais.

requerida quanto à perfectibilização de um plano em linha de telefonia móvel da autora diversa daquela [em relação à qual a consumidora pretendia fosse efetuado o plano], assim como, não se constata a verossimilhança da alegação da mesma no [que se refere ao fato] [de que possuía duas linhas de telefonia móvel, tampouco junto à empresa requerida].

Os três exemplos acima são mais uma amostra de períodos longos com apenas 1 frase com vários constituintes de hierarquia, fator considerado dificultador para a compreensão leitora. No exemplo (7.15), para ilustrar, temos uma oração reduzida de infinitivo atuando como sujeito do verbo “cabendo”. Essa oração é composta de 8 palavras, com uma sentença encaixada qualificando o substantivo “tratamento”. Por estar no final da oração, esse encaixe não quebra a linearidade do processamento na leitura e, por isso, em princípio, não seria um fator de maior complexidade. Entretanto, após a oração assinalada, há ainda duas orações relativas encaixadas qualificando elementos das orações anteriores.

Prosseguindo em nossa descrição, adotamos o Procedimento 2, a seguir especificado, no qual verificamos a forma como os sujeitos oracionais, os sujeitos pronominais e os sujeitos elípticos se apresentavam.

Procedimento 2:

Nesta etapa, verificamos a forma como se apresentavam os sujeitos oracionais, os sujeitos pronominais e os sujeitos elípticos.

a) Sujeitos Oracionais

A Tabela 7.10 mostra como se apresentam os sujeitos oracionais: se na forma de orações desenvolvidas, com a presença da conjunção, ou se reduzidas, em uma das formas nominais do verbo.

Tabela 7.10 – Tipos de CFSs Oracionais

TIPOS DE SUJEITOS ORACIONAIS	NÚMERO DE OCORRÊNCIAS	PERCENTUAL DE OCORRÊNCIAS
Oração desenvolvida (com a conjunção)	265	54,52%
Oração reduzida	221	45,48%
TOTAL	486	100%

Como podemos ver, pela Tabela 7.10, os sujeitos oracionais são representados, em maior parte, por orações que apresentam o conectivo exposto. Mas a diferença para

as ocorrências relativas às orações reduzidas é pouca, aproximadamente 9%. As orações reduzidas, por apresentarem o verbo no infinitivo em sua maioria sem flexão (apenas 2 situações de infinitivo flexionado) podem ser um fator dificultador para a leitura, visto que não se pode relacionar um elemento textual ou situacional a ele. Seriam casos de sujeitos indeterminados.

b) Sujeitos Pronominais

Já mencionamos que, nas pesquisas sobre leitura e letramento, os pronomes são também considerados um fator dificultador da compreensão leitora. Assim se dá porque, em muitas situações, as referências pronominais podem se referir a mais de um elemento expresso no texto, e, na busca do referente pronominal, a fluência da leitura tende a ficar comprometida. É o que ocorre com o pronome “ele”, no exemplo (7.2), repetido aqui.

(7.2) A testemunha Fulano de Tal, em seu depoimento, declarou que o autor procurou o local onde ele trabalhava para fazer passaporte, visto, bem como auxílio na compra do bilhete de viagem. Afirmou ao requerente que não tinha experiência naquilo, que nunca havia comprado passagem de avião.

O pronome “ele” no exemplo (7.2) pode se referir à “testemunha Fulano de Tal” como também a “o autor”. O leitor terá de procurar as pistas contextuais para compreender que este “ele” só pode ser a “testemunha Fulano de Tal”, pois, em princípio, “o autor” não iria até o seu local de trabalho apenas para fazer o passaporte. Sublinhamos também o pronome “naquilo”, cujo referente não está muito preciso no trecho.

A Tabela 7.11, a seguir, mostra os tipos de sujeitos pronominais presentes em nossa amostra com suas respectivas quantidades.

Tabela 7.11 – Tipos de CFSs Pronominais

TIPOS DE PRONOMES	NÚMERO DE OCORRÊNCIAS	PERCENTUAL DE OCORRÊNCIAS
Relativos (que, quem, qual)	1020	69%
Demonstrativos	230	16%
Retos	38	3%
Mesmo ⁹⁶	103	7%
Indefinidos	68	5%
TOTAL	1459	100%

⁹⁶ Colocamos “mesmo” separadamente, porque atua diversamente na amostra – ora como pronome reto, ora como pronome demonstrativo.

Pela Tabela 7.11, verificamos que os pronomes relativos (que, quem e qual) ocupam o primeiro lugar entre os sujeitos pronominais. Juntos, num total de 1020 ocorrências, representam um percentual de 69% do total desse conjunto de sujeitos. Além disso, essas 1020 ocorrências de pronomes relativos na função de sujeito indicam a existência de 1020 orações relativas encaixadas, o que representa um percentual de 8,7% do total de orações da amostra. As orações encaixadas, como já mencionamos, agregam informação aos constituintes da oração anterior, o que demanda maior custo de processamento na leitura e, em princípio, oferece maior dificuldade para a compreensão leitora.

A seguir, apresentamos alguns exemplos de sujeitos pronominais da nossa amostra:

- (7.18) O orçamento da folha 61 emitido por Fulano de Tal, [**que individualiza as peças**], está incompleto visto que não contempla todos os itens avariados e, por isso, não pode ser levado em consideração.
- (7.19) Assim, configurado o dano moral *in re ipsa*, [**que prescinde de prova da ocorrência de prejuízo concreto**], [**o qual se presume**], conforme as regras de experiência comum é devida a reparação extrapatrimonial pleiteada.
- (7.20) Ainda que a parte autora atribua ao requerido XXX a imagem de um rapaz de moletom preto e boné, **que** pega um paralelepípedo do vidro traseiro, já danificado, e atira para o interior do veículo pelo lado direito, já sem vidros, **isto** em vídeo gravado por populares e juntado pela autora nos presentes autos, não há como precisar que seja o requerido, sendo que **o mesmo** rechaça ser **ele** tal personagem. **Isto** consta no vídeo em que o agente de trânsito está em pé sobre o veículo e diz que já prendeu o condutor do veículo.

No exemplo (7.18), temos o pronome “que” atuando como sujeito e retomando o substantivo “orçamento” da oração anterior. Porém, seria possível também pensar que o referente deste “que” seja “Fulano de Tal”. Para o leitor ter certeza de que o referente é “orçamento”, e não “Fulano de Tal”, ele precisa prosseguir na leitura e verificar que todos os demais verbos só podem estar relacionados a “orçamento”.

No exemplo (7.19), há dois pronomes relativos com função de sujeito. A segunda ocorrência inicia uma oração encaixada dentro da anterior, que também é encaixada. Trata-se de mais um exemplo de um período com vários constituintes hierarquicamente

diferentes, o que, como já mencionamos também, é fator que dificulta a compreensão leitora.

O exemplo (7.20), representado por um período composto de duas frases, contém 5 ocorrências de sujeitos pronominais. As primeiras 4 ocorrências estão na primeira frase, que, com 72 palavras organizadas por meio de vários constituintes hierarquicamente diferentes, é mais um exemplo desse tipo de estrutura frasal em nossa amostra.

Por último em nossa lista de sujeitos, estão os elípticos. Na próxima seção, veremos como ele se apresenta em nossa amostra.

c) **Sujeitos Elípticos**

Na seção 7.1.1, já esclarecemos que, apesar de nossa análise centrar-se no constituinte expreso da oração, computamos os sujeitos elípticos também, porque, mesmo não havendo um elemento explícito ocupando o lugar do sujeito, há um referente, dentro ou fora do texto, que exerce essa função. A busca por tal referente, como também já mencionamos, é mais custosa e sobrecarrega a memória de trabalho do leitor, sendo, em princípio, mais complexa do que a identificação de um sujeito explícito na oração.

As formas verbais relativas aos sujeitos elípticos de nossa amostra apresentam-se conforme consta na Tabela 7.12, a seguir.

Tabela 7.12 – Formas Verbais Referentes aos CFSs Elípticos

FORMAS VERBAIS	NÚMERO DE OCORRÊNCIAS	PERCENTUAL DE OCORRÊNCIAS
Verbos na 3ª pessoa	1734	42%
Verbos na 1ª pessoa	664	16%
Verbos no Infinitivo	887	21%
Verbos no Gerúndio	592	14%
Verbos no Particípio	282	7%
TOTAL	4159	100

Pela Tabela 7.12, verificamos que os sujeitos elípticos com os verbos na 3ª pessoa são os de maior frequência em nossa amostra. Dentro desse grupo, 1.577 verbos estão na 3ª pessoa do singular e 157 estão na 3ª pessoa do plural. Segundo Vilela (1988, p. 27), as “formas verbais de terceira pessoa são, portanto, potencialmente ambíguas, já que podem se referir, em princípio, a qualquer sintagma nominal expreso no discurso precedente”.

A seguir dois exemplos desse tipo de ocorrência:

- (7.21) Aduziu a Autora que somente **adquiriu** os móveis e utensílios da Requerida porque **pretendia** dar continuidade ao negócio. **Informou**, ainda, que **teve** prejuízos eis que **acreditava** que a Requerida possuía alvará de licença sanitária para funcionar como lancheria.
- (7.22) Aduziram as autoras que **organizaram** com antecedência as férias, fazendo pesquisas em diversas companhias de viagem e empresas de turismo, para que **tivessem** uma virada de ano em um hotel no Uruguai e **desfrutassem** dos dias de descanso ao lado de seus esposos. Assim sendo, **adquiriram** junto à companhia de viagem e turismo requerida um pacote de viagem para cidade de Punta Del Este – Uruguai [...].

Nos dois exemplos acima, não parece difícil encontrar o referente, no próprio texto, para os verbos destacados: “a Autora”, no primeiro, e “as autoras”, no segundo. De todo modo, é preciso considerar a potencial ambiguidade dos verbos na 3ª pessoa.

Em relação aos verbos flexionados na 1ª pessoa, retomamos aqui que a grande maioria deles se refere ao redator do texto, e que essa informação é possível de ser depreendida do contexto situacional.

As outras formas verbais relativas a sujeitos elípticos são nominais: os verbos no infinitivo estão em primeiro, seguidos dos no gerúndio, vindo por último os participípios. Abaixo apresentamos mais alguns exemplos de sujeitos elípticos extraídos de nossa amostra, com algumas situações em que os verbos estão na forma nominal.

- (7.23) **Citada** (folha 83), a parte Requerida contestou o feito (folhas 81/104). **Alegou** que o poste da rede de energia foi instalado no local há anos [...]
- (7.24) A parte Ré realizou vistoria no local, **exigindo** o montante de R\$ 5.859,51 (cinco mil oitocentos e cinquenta e nove reais e cinquenta e um centavos) para **prestar** o serviço.

No exemplo (7.23), a forma verbal “citada”, num processo catafórico, refere-se ao sujeito “a parte requerida”. E a forma verbal “alegou” retoma este mesmo sujeito por um processo anafórico. Segundo Fulgêncio e Liberato (2010), a catáfora implica uma sobrecarga na memória de curto prazo, porque exige que o item reduzido tenha de ser guardado na memória até que seu referente apareça no texto. Isso, conseqüentemente, acarreta maior trabalho para o leitor e maior dificuldade na leitura.

E, no exemplo (7.24), as formas verbais “exigindo” e “prestar” retomam, anaforicamente, o sujeito “a parte Ré”.

Ainda segundo as autoras Fulgêncio e Liberato (2010), a busca de um referente na anáfora pode prejudicar a legibilidade do texto. Se essa relação for transparente – os referentes sendo facilmente identificados -, a leitura segue sem problemas, como parece

ser o caso do exemplo (24). Porém, havendo dificuldade de encontrar o referente, “a leitura pode ser atrasada ou até mesmo interrompida” (FULGÊNCIO; LIBERATO, 2010, p. 85). É o que ocorre no exemplo (7.25), a seguir.

(7.25) Restou comprovado nos autos a culpa exclusiva do demandado, que invadiu a via preferencial pela qual a demandante trafegava, **desrespeitando** a sinalização e a preferencialidade da via [...]

O gerúndio “desrespeitando”, sublinhado no exemplo (7.25), pode, numa leitura inicial, ser atribuído à “demandante”, que consta na oração imediatamente anterior. Todavia, essa interpretação é incoerente com o contexto. Tal processo exige, como já observamos, mais trabalho para o leitor e, conseqüentemente, pode ser considerado um fator de complexidade leitora.

Antes de finalizarmos as observações sobre os sujeitos elípticos de nossa amostra, julgamos importante retomar aqui uma pesquisa realizada por Finatto *et al.* (2011), já por nós mencionada nas considerações finais do Estudo-Piloto 1. Nesse estudo, as autoras compararam, entre si, os jornais Zero Hora e Diário Gaúcho. A comparação baseou-se em várias métricas do sistema Coh-Metrix, inclusive quanto à presença de elipses. Um dos pressupostos do estudo foi o de que o jornal Zero Hora, dirigido para pessoas de maior escolaridade, teria uma linguagem mais complexa do que a do Diário Gaúcho. Porém, quanto às elipses, verificaram que seu uso, nos textos, não é um fator diferencial entre esses dois tipos de jornais. A conclusão foi, assim, a de que a presença de elipses não seria um fator decisivo na avaliação da complexidade textual naquele cenário discursivo.

De todo modo, os sujeitos elípticos, como já vimos, exigem mais da memória de trabalho do leitor e eles representam os sujeitos mais frequentes em nossa amostra. Diante disso, a análise de como se comportam textos-sentenças de nossa amostra quanto a esse aspecto parece-nos bastante importante para a aferição do quanto o fenômeno dos sujeitos elípticos influi para a acessibilidade textual de nossas Sentenças, no que se refere a um leitor com pouca escolaridade e pouca experiência em leitura.

Após analisarmos os CFSs em nossa Amostra 1, passamos à análise dos constituintes frasais com função de complementos verbais, doravante CFCVs. Como já referimos, estudos sobre a complexidade textual informam que esses elementos são variáveis importantes na aferição da complexidade textual e, mais especificamente sentencial – frasal e oracional.

7.1.2.2 Constituintes Frasais com Função de Complementos Verbais (CFCVs)

Para a extração dos CFCVs, lemos novamente cada uma das orações presentes em nossa Amostra 1. Esse procedimento gerou um total de 10.840 CFCVs. Como nosso foco de análise não está no aspecto semântico, mas na forma como se apresentam esses elementos, distribuímos os CFCVs em três grandes grupos:

- a) complementos verbais preposicionados (CFCVPs);
- b) complementos verbais não preposicionados (CFCVNPs); e
- c) complementos verbais representados por pronomes (CFCVPRs).

Os CFCVPs estão relacionados ao verbo por meio de uma preposição, e os CFCVNPs complementam o verbo sem a intermediação de preposição. Os CFCVPRs, foram colocados à parte, mesmo que alguns possam vir acompanhados de preposição. Assim fizemos, porque normalmente os CFCVPRs são representados por uma ou duas palavras e muitos introduzem orações encaixadas. E as orações encaixadas merecem uma análise à parte em razão da maior complexidade que podem trazer aos textos em que se inserem.

Entre os CFCVPs, encontram-se complementos qualificados pela Gramática Tradicional como objeto indireto, predicativo, complemento circunstancial e agente da passiva. Entre os CFCVNPs, encontram-se complementos verbais classificados como objeto direto e predicativo. E entre os CFCVPR também encontramos essas categorias de complementos verbais.

Na Tabela 7.13, apresentamos a distribuição quantitativa desses CFCVs em nossa Amostra 1.

Tabela 7.13 – Constituintes Frasais com Função de Complemento Verbal (CFCVs)

TIPOS	OCORRÊNCIAS
PREPOSICIONADO - CFCVP	3108
NÃO PREPOSICIONADO - CFCVNP	6998
PRONOMINAIS - CFCVPR	734
TOTAL	10.840

Pelos dados da Tabela 7.13, verificamos inicialmente que o número de CFCVs que se ligam diretamente ao verbo é mais que o dobro do que aqueles que se relacionam ao verbo por meio da preposição. A ausência da preposição é uma palavra a menos na extensão desses constituintes, o que pode ser um facilitador no processamento da leitura.

Os CFCVPRs, por sua vez, representam apenas 6,77% do total de CFCVs da Amostra 1. Como os pronomes precisam de um referente para serem compreendidos, a pouca presença deles como complementos verbais também é indicativo de menor complexidade textual quanto a esse aspecto.

Após a distribuição apresentada na Tabela 7.13, subdividimos os CFCVs, tanto preposicionados como não preposicionados, de acordo com a estrutura em que se apresentam: por sintagma nominal ou por sintagma oracional. Os CFCVPRs foram subdivididos conforme o tipo de pronome.

A Tabela 7.14 traz os dados dessa subdivisão.

Tabela 7.14 – CVFCVs e seus Subtipos Presentes na Amostra 1

TIPO	SUBTIPO	OCORRÊNCIAS
CFCVP	SINTAGMAS NOMINAIS	2983
	SINTAGMAS ORACIONAIS	125
CFCVNP	SINTAGMAS NOMINAIS	5864
	SINTAGMAS ORACIONAIS	1134
CFCVPR	relativo	185
	oblíquo (lhe, me, o, a, se)	533
	reto	16
TOTAL		10840

A seguir, apresentamos um trecho, exemplo (7.26), com alguns CFCVs assinalados conforme a classificação acima.

(7.26) Artigo 34. O condutor que queira executar (a) uma manobra deverá certificar-se (c) (b) de que pode executá-la (c) sem perigo para os demais usuários da via que o (c) seguem, precedem ou vão cruzar com ele (c), considerando sua posição, sua direção e sua velocidade.

Em (a), temos um CFCV sintagma nominal não preposicionado – *uma manobra*.

Em (b), temos um CFCV sintagma oracional preposicionado – *de que pode executá-la sem perigo para os demais usuários da via que o seguem, precedem ou vão cruzar com ele*.

Em (c), temos CFCVs pronominais – *se, la, e com ele*

Entre os subtipos de CFCVs apresentados na Tabela 7.14, verificamos que 8.847 são representados por sintagmas nominais contra 1259 ocorrências de sintagmas oracionais. As orações, em princípio, podem apresentar maior número de palavras do que os sintagmas nominais. Também podem oferecer maior dificuldade para o processamento da leitura, uma vez que sua estrutura é mais complexa que a de um sintagma nominal. Tanto é assim que, conforme já mencionamos, são denominados por Perini (1994, 204) como *sintagmas nominais complexos*.

Ainda em relação aos CFCVs oracionais, verificamos quantos eram representados por orações desenvolvidas e quantos por orações reduzidas. É o que apresentamos na Tabela 7.15, abaixo:

Tabela 7.15 – Tipos de CFCVs Oracionais

TIPOS DE COMPLEMENTOS VERBAIS ORACIONAIS	NÚMERO DE OCORRÊNCIAS	PERCENTUAL DE OCORRÊNCIAS
Oração desenvolvida (com a conjunção)	967	76,80%
Oração reduzida	292	23,20%
TOTAL	1259	100%

Como podemos ver, pela Tabela 7.15, tal como ocorreu com os sujeitos oracionais, os CFCVs são representados, em maior parte, por orações que apresentam o conectivo expresso. E a proporção dessa ocorrência foi quase três vezes maior que a ocorrência de CFCVs de orações reduzidas. Esse fato, em princípio, é um facilitador para a leitura, pois as orações desenvolvidas, apesar de serem mais complexas que os sintagmas nominais, apresentam o conectivo que as liga à outra oração e também os demais elementos constituintes da oração. As reduzidas, por apresentarem o verbo numa forma nominal, podem ser um fator dificultador para a leitura, visto que pode ser mais difícil encontrar o elemento textual ou situacional que ao verbo se refere.

Diante desse dado, em princípio, os CVCFs da Amostra 1 apresentam-se, em sua grande maioria, numa forma que facilita mais o processamento da leitura.

Abaixo, colocamos exemplos de CVCFs oracionais da Amostra 1.

- (7.27) Ressalto, inicialmente, que a responsabilidade da empresa ré no caso em tela é objetiva, independentemente de prova de culpa, nos termos do artigo 37, § 6º, da Constituição Federal e do artigo 14 do Código de Defesa do Consumidor [...].
- (7.28) Tal legislação estabelece serem direitos básicos do consumidor, entre outros, a facilitação de defesa de seus direitos, inclusive com a inversão do ônus da prova, a seu favor, no processo civil, quando, a critério do juiz, for verossímil a alegação ou quando for ele hipossuficiente, segundo as regras ordinárias de experiência (artigo 6, inciso VIII, CDC).

Os trechos sublinhados nos exemplos (7.27) e (7.28) são orações atuando como complemento verbais. O primeiro, (7.27), apresenta um CFCV de oração desenvolvida composta por 37 palavras. E o segundo, (7.28), apresenta uma oração reduzida de 53 palavras. Chama a atenção a extensão de cada um desses CFCV, fato que retomaremos nos exemplos colocados na seção relativa ao Procedimento 2 em nossas análises.

Vejamos, agora, o que ocorre com os CFCVPRs.

Em relação aos CFCVPRs, os pronomes que mais ocorreram foram os oblíquos, cuja função essencial é atuar como complemento. No exemplo (7.26), apresentamos alguns. Também encontramos alguns poucos casos de pronomes retos na função de complemento. É o que ocorre no exemplo (7.26), onde se vê o pronome **ele** antecedido da preposição **com** de modo a complementar o verbo **cruzar**.

A segunda colocação no *ranking* dos CFCVPRs é referente aos pronomes relativos. Os pronomes relativos podem oferecer alguma dificuldade para o processamento da leitura, porque muitas vezes apresentam problemas relacionados à anáfora. Além disso, eles sempre introduzem orações encaixadas, que, por sua vez, quando malconstruídas ou empregadas em excesso, também podem dificultar o processamento da leitura.

A seguir relacionamos alguns exemplos de pronomes relativos como CFCVs.

- (7.29) Enfatizou que a colisão ocorreu em razão da alta velocidade [que o Autor mantinha no momento do fato].

Neste exemplo (7.29), o pronome **que** complementa, sem preposição, o verbo **mantinha**, retomando **alta velocidade** da oração anterior.

(7.30) XXXX Propaganda Limitada ME ajuizou a presente ação de reparação de danos contra o Banco YYYYYYYY S/A, asseverando, em síntese, que em 03.08.2017 teve o seu crédito negado, pois quando se submeteu à análise de crédito foi surpreendida com uma negativa, em face de existir uma inscrição negativa nos cadastros do SERASA, em razão do apontamento de registro de um lançamento efetuado pelo demandado, com vencimento datado de 15.10.2015, no valor de R\$ 11.137,99, **cuja** dívida a demandante não reconhece.

No exemplo (7.30), encontramos o pronome **cujo**, de uso mais raro. Ele introduz a última oração de uma frase com 81 palavras, mas não é o núcleo do CFCV em questão. Porém, é mais um exemplo a ilustrar a tendência de frases longas, com excesso de elementos hierarquizados.

Tendo um panorama de como se apresentam os CFCVs da Amostra 1, seguimos nossas análises, adotando um segundo procedimento, tal qual fizemos com os CFSs: passamos a analisar a extensão dos CVCFs da Amostra 1.

Procedimento 2:

Após a distribuição por grupos de CFCVs, verificamos a extensão desses elementos, ou seja, de quantas palavras são constituídos. Os CFCVPRs não foram contabilizados nessa distribuição.

A Tabela 7.16 mostra essa distribuição.

Tabela 7.16 – Ocorrência de CFCVs por Tipo e Número de Palavras

TIPO	SUBTIPO	NÚMERO DE PALAVRAS						total
		1-5	6-10	11-20	21-30	31-40	41+	
CFCVP	SINTAGMAS NOMINAIS	1926	562	333	91	34	37	2983
	SINTAGMAS ORACIONAIS	9	22	36	22	22	14	125
CFCVNP	SINTAGMAS NOMINAIS	3666	1260	697	155	60	26	5864
	SINTAGMAS ORACIONAIS	83	258	410	225	93	65	1134
TOTAL		5684	2102	1476	493	209	142	10106

Como se vê na Tabela 7.16, mais de 50% dos CFCVs em forma de sintagmas nominais apresentam de 1 a 5 palavras, seguidos, em termos de ocorrências de complementos na faixa de 6 a 10 palavras, e diminuindo a ocorrência à medida que o número de palavras aumenta. Com os CFCVs em forma de sintagmas oracionais, a ocorrência maior se dá entre as faixas de 6 a 30 palavras.

Abaixo apresentamos alguns exemplos de CFCVs encontrados em nossa amostra.

(7.31) A finalidade da reparação do dano moral é oferecer **compensação ao lesado** [..].

No exemplo (31), temos dois CFCVs em forma de sintagma nominal: **compensação** (1 palavra; ligado ao verbo sem preposição) e **ao lesado** (2 palavras, uma delas a preposição que o relaciona ao verbo).

(7.32) É certo, porém, que tal quantia deve ser **suficiente para reparar o dano, e nunca para causar locupletamento sem causa ao lesado.**

No exemplo (7.32), o CFCV sublinhado é composto por 14 palavras, porque o adjetivo suficiente é complementado por duas orações:

- para reparar o dano; e
- nunca para causar locupletamento sem causa ao lesado.

Essas orações, por sua vez, também apresentam complementos verbais:

- o dano (na primeira, 2 palavras);
- locupletamento sem causa e ao lesado (na segunda, 3 palavras e 2 palavras).

(7.33) Ainda, a autora procurou solucionar a celeuma administrativamente, através do Procon, sem sucesso, comprovando **a desídia da ré perante a autora, consumidora final do serviço prestado.**

No exemplo (7.33), o trecho em negrito do CFCV sublinhado qualifica a palavra **autora**, integrante desse complemento, e por isso é computado na totalidade desse CFCV. Temos, neste caso, um CFCV em forma de sintagma nominal com 12 palavras e relacionado ao verbo sem preposição.

(7.34) Observo, também, que a Carta AR informando nova data de audiência teve **retorno negativo com certidão de que a entrega não pode ser efetuada pois o cliente mudou-se (folha 137)**, o que demonstra o descumprimento de obrigação processual do requerido.

A oração em negrito no exemplo (7.34) foi incluída no CFCV que inicia em **retorno negativo**, pois expressa uma circunstância relativa à oração **que a entrega não pode ser efetuada**. Esta oração, por sua vez, está subordinada à **certidão**, que compõe esse constituinte.

Muitas situações como essas ocorrem entre os CFCVs encontrados em nosso CE. Abaixo trazemos mais alguns exemplos.

(7.35) Alegou que o corretor da Bradesco Saúde informou que não seria necessário que a empresa autora providenciasse o cancelamento do plano antigo, **já que essa medida seria tomada pela nova operadora.**

(7.36) Todavia, saliento, em primeiro lugar, que não houve informação, a este Juízo, sobre o efetivo deferimento da recuperação judicial, tampouco do plano respectivo, com a inclusão dos credores destes autos naquele rol, **de modo a se saber quais os prazos de cumprimento das obrigações assumidas, mormente aquelas decorrentes de demandas judiciais oriundas de danos ao consumidor.**

Nos exemplos (7.35) e (7.36), temos CFCVs oracionais com circunstâncias expressas por meio de orações adverbiais (em negrito) que foram computadas na totalidade desses CFCVs. Por essa razão, tornam-se bem longos, o primeiro com 30 palavras, e o segundo com 44.

Situações como essas dos exemplos (7.35) e (7.36) confirmam a tendência do emprego de frases longas nos textos jurídicos. É o que se vê nos exemplos abaixo.

(7.37) A controvérsia dos autos tem por cerne à (sic) discussão sobre a restituição dos valores pagos a título de taxa de corretagem realizado (sic) pela autora aos corretores Senhores FULANO DE TAL, BELTRANO DE TAL e SICRANO DE TAL, em decorrência de contrato de promessa de compra de um imóvel realizado entre a autora e a corré XXXXXXXXX, em que havia previsão da existência do pagamento dessa comissão.

O núcleo desse CFCV de 59 palavras é **discussão**. Tal substantivo é especificado com o que vem após **sobre**, que é representado por uma sucessão de itens subordinados entre si. Por exemplo:

- **pagos a título de taxa de corretagem** [...] está subordinado a **valores**;
- **valores**, por sua vez, está subordinado a **restituição**.

(7.38) O Tribunal a quo negou provimento ao apelo interposto pelo ora agravante, sob o fundamento de que, nas relações de consumo, as cláusulas limitativas de direito serão sempre interpretadas a favor do consumidor, desse modo, ao assim decidir, adotou posicionamento consentâneo com a jurisprudência desta egrégia Corte, que se orienta no sentido de considerar que, em se tratando de contrato de adesão submetido às regras do CDC, a interpretação de suas cláusulas deve ser feita da maneira mais favorável ao consumidor, bem como devem ser consideradas abusivas as cláusulas que visam a restringir procedimentos médicos.

O exemplo (7.38) é composto por uma frase com 96 palavras. Poderia, sem maiores alterações, ser colocado um ponto final após **consumidor** e antes de **desse modo**. Tal procedimento diminuiria a extensão da frase, facilitando a sua leitura e compreensão.

Além disso, o trecho sublinhado após **adotou** representa um CFCV com 56 palavras organizadas também numa sucessão de elementos hierarquicamente subordinados. Essa sucessão de encadeamentos numa mesma frase pode gerar algumas ambiguidades, como a que se dá com o pronome **que** após **Corte**. É possível encontrar três referentes para ele: *posicionamento*, *jurisprudência* ou *Corte*.

(7.39) As partes demandadas, no mérito, disseram que os prejuízos materiais suportados pela autora, (sic) ocorreram única e exclusivamente por sua culpa, que optou pelo desfazimento da compra e venda após a assinatura do contrato, tendo elas atuado com êxito na intermediação entre a compradora e o vendedor, tanto que o contrato se deu nos termos avençados com a demandante, ou seja, o apartamento sendo adquirido pelo valor de 300 mil e sendo pago com a dação do imóvel da autora no valor de 200 mil, mais 60 mil em dinheiro e 40 mil financiados.

O exemplo (7.39) contém 94 palavras, e 87 delas compõem um CFCV oracional iniciado com o **que** após **disseram**. É mais um exemplo de constituinte com várias ideias sucessivamente encaixadas, dificultando a sua leitura e, conseqüentemente, a sua compreensão.

Concluindo, cremos que os dados encontrados nas análises realizadas neste Estudo-Piloto comprovam o que dissemos ao justificar nossa pesquisa: no que se refere à linguagem empregada nas Sentenças dos JECs, parece que nem sempre ela atende aos princípios da art. 2º da Lei 9.099/95, entre eles os da simplicidade e da informalidade.

Afinal, a simplicidade na linguagem, como vimos até aqui, não combina com períodos longos de 1 frase só com vários níveis de subordinação. Não combina com um percentual de 63% de sujeitos (elípticos, pós-verbais e pronominais) que demandam maior sobrecarga na memória de trabalho do leitor com. Não combina também com complementos verbais com 87 palavras igualmente relacionadas por diversos níveis de subordinação. Além disso, como já vimos em outros estudos-pilotos, antes citados, o uso de palavras como ADUZIR e de outras características, como quantidade de negações e de pronomes, só fazem aumentar essa nossa percepção de complexidade.

De todo modo, para verificarmos em que medida essas características são referentes apenas aos textos do nosso CE, julgamos importante pesquisar como se apresentam esses constituintes frasais nos jornais populares que fazem parte dos nossos *corpora* de contraste: Jornal Diário Gaúcho, de Porto Alegre; e Jornal Massa, de Salvador.

Como já referimos, tais jornais são considerados populares por terem por objetivo oferecer notícias e informação de entretenimento e serviço às classes C, D e E. Para tanto, valem-se de muitas imagens, textos curtos e procuram usar uma linguagem em princípio mais simples e mais próxima da que é empregada no cotidiano do povo brasileiro.

Acreditamos que, ao compararmos o comportamento dos constituintes frasais com função de sujeito e de complementos verbais das sentenças com os dos jornais populares, teremos uma melhor dimensão do quanto o emprego desses elementos nas sentenças dos JECs se aproxima ou se afasta de uma linguagem mais próxima de um PPE, que comporia os textos desses jornais populares.

Destinamos, então, a próxima seção para a apresentação deste estudo contrastivo.

7.2 ESTUDO-PILOTO 4 – CONSTITUINTES FRASAIS EM AMOSTRAS DO CE, DO DG E DO JM

Este Estudo-Piloto apresenta uma análise contrastiva entre uma subamostra de sentenças do nosso CE (Amostra 2) com amostras de textos de dois de nossos *corpora* de contraste: o Jornal Diário Gaúcho (DG) e o Jornal Massa (JM). Como não teríamos tempo hábil para analisar a mesma proporção de textos (110 sentenças) nos dois jornais de contraste, optamos por gerar uma amostra de cada um desses conjuntos de textos. Tal decisão baseou-se na representatividade que a amostra das sentenças (Amostra 1) obteve em relação à totalidade do CE, tal como apresentamos na Tabela 7.1 do Estudo-Piloto 3.

A partir daqui, no item 7.2.1, explicamos como procedemos para a compilação das amostras deste Estudo-Piloto. E, no item 7.2.2, comentaremos os dados das análises realizadas.

7.2.1 Constituição das amostras

Compilamos, inicialmente, uma amostra (Amostra 2) da primeira amostra das sentenças do nosso CE (Amostra 1). A Tabela 7.17 apresenta alguns dados das duas amostras.

Tabela 7.17 – Dados das Amostras 1 e 2 do CE

	AMOSTRA 1	AMOSTRA 2	Percentual da Amostra 2 em relação à Amostra 1
TOTAL DE SENTENÇAS	110	12	10,90%
TOTAL DE FRASES	7.491	760	10,14%
TOTAL DE TOKENS	137.700	13.518	9,83%
TOTAL DE TYPES	8.477	2.504	
PAL/FRASE	19,25	18,67	

Conforme se vê na coluna relativa a percentuais na Tabela 7.17, a Amostra 2 representa em torno de 10% da Amostra 1 em relação ao número de sentenças, ao número de frases e ao número de *tokens*. O número de *types* não obedece a essa proporção, porque a contagem dos *types* não tem o mesmo sentido no todo e na parte. Por exemplo, um *type* que ocorre em duas partes de um *corpus* conta por 1 na totalidade, mas por 2 nas partes. Como são duas partes, em cada uma aquela *type* vai contar como 1 elemento. Em relação ao número de palavras por frase, as amostras são muito próximas: a Amostra 1 apresenta 19,25 palavras por frase, enquanto a Amostra 2 apresenta 18,67.

A seleção das 12 sentenças procurou também contemplar uma proporção semelhante em relação à temática, como se vê na Tabela 7.18, a seguir:

Tabela 7.18 – Total de Sentenças por Temática nas Amostras 1 e 2

TEMÁTICA SENTENÇA	TOTAL DE SENTENÇAS DA AMOSTRA 1	TOTAL DE SENTENÇAS DA AMOSTRA 2
Água	12	1
Acidente de trânsito	6	1
Energia Elétrica	17	2
Inadimplência	13	2
Plano de Saúde	17	2
Prestação de Serviços	25	2
Telefonia	20	2
TOTAL	110	12

Selecionamos duas sentenças das temáticas em que havia mais textos na Amostra 1 (Energia, Inadimplência, Plano de Saúde, Prestação de Serviços e Telefonia). Nas temáticas com menor quantidade (Água e Acidente de Trânsito), selecionamos 1 sentença apenas. Nas temáticas com duas sentenças na Amostra 2, uma foi do ano de 2018 e outra de 2019.

Para a compilação das amostras dos *corpora* dos Jornais DG e JM, selecionamos, do *corpus* total de cada um deles, um conjunto de matérias (textos) que se aproximasse, em número de *tokens* da Amostra 2 das Sentenças. Com isso, para este estudo comparativo, que faz um contraponto entre a sintaxe da linguagem jurídica e a linguagem do jornalismo popular, ficamos com os conjuntos de textos relacionados na Tabela 7.19.

Tabela 7.19 – Dados das Amostras 2, do DG e do JM

	AMOSTRA 2	DIÁRIO GAÚCHO (DG)	JORNAL MASSA (JM)
TOTAL DE TOKENS	13518	13712	13623
TOTAL DE TYPES	2504	3712	3790
TOTAL DE FRASES	760	1057	822
PAL/FRASE	18,67	14,36	16,90

Como se vê na Tabela 7.19, os três *corpora* são muito próximos em relação ao número de *tokens*, porém a amostra de Sentenças ficou com um número de *types* inferior ao número de *types* das amostras dos jornais. Este fenômeno significa menor variação lexical e pode ser atribuído ao fato de as sentenças apresentarem uma linguagem especializada, que, em princípio, contém muita terminologia e repetição de palavras.

Também podemos verificar, na Tabela 7.19, que as sentenças apresentam um menor número de frases e maior número de palavras por frase que as amostras dos jornais. Este dado pode indicar que os CFSs e os CFCVs apresentem um maior número de palavras que estes mesmos constituintes nas amostras dos jornais. Veremos se isso se confirma.

7.2.2 Processamento e Análise dos Dados

Com as três amostras compiladas, extraímos os dados referentes aos sujeitos e complementos verbais das sentenças da Amostra 2, que foram armazenados durante a análise da Amostra 1. Após essa etapa, tal qual fizemos com a análise da Amostra 1, lemos cada uma das orações presentes nas amostras do DG e do JM e listamos todos os CFSs e todos os CFCVs.

Iniciaremos a descrição de nossas análises pelos CFSs, na seção 7.2.2.1 e, posteriormente, na seção 7.2.2.2, apresentamos os dados relativos aos CFCVs.

7.2.2.1 Constituintes Frasais com Função de Sujeito (CFSs)

Os dados relativos aos CFSs obedeceram aos mesmos critérios estabelecidos para esses constituintes no Estudo-Piloto 3. Ou seja, os CFSs foram classificados, inicialmente, em sujeitos elípticos e realizados. Entre os realizados, foram classificados em: pré-verbal (sujeitos colocados antes do verbo); pós-verbal (sujeitos colocados após o verbo); pronominal (sujeitos representados por pronomes – relativos, demonstrativos, retos etc.); e oracional (oração como sujeito de outro verbo).

A classificação quanto à posição do sujeito – pré ou pós-verbal – deu-se apenas em relação aos CFSs constituídos por sintagmas nominais.

Na Tabela 7.20, temos o total de CFSs selecionados com a quantidade encontrada para cada tipo.

Tabela 7.20 – Ocorrências por Tipos de CFSs das Amostra 1, DG e JM

<i>CORPORA</i>	TIPOS DE SUJEITOS					
	ELÍPTICOS	REALIZADOS				TOTAL
		SINTAGMA NOMINAL		PRONOMINAL	ORACIONAL	
		PRÉ	PÓS			
AMOSTRA 1	460	377	200	96	45	1178
DG	457	638	97	202	12	1406
JM	464	680	92	191	14	1441
TOTAL	1381	1695	389	489	71	4025

Os dados da Tabela 7.20 mostram que o total de sujeitos das sentenças é menor que o total de sujeitos das amostras dos jornais. Esse dado é compatível com o fato de as sentenças apresentarem um menor número de frases que os jornais, conforme visto na Tabela 7.19.

A Tabela 7.20 também mostra que a ordem de ocorrência por tipo de CFSs na Amostra 2 foi idêntica à da Amostra 1: em primeiro lugar, os sujeitos elípticos, depois os pré-verbais, seguidos dos pós-verbais, pronominais e oracionais. Mais um dado a comprovar a representatividade das amostras.

A seguir, apresentamos algumas observações sobre cada tipo de sujeito da Tabela 7.20.

a) **Sujeitos Elípticos**

Os dados relativos aos sujeitos elípticos chamam a atenção na Tabela 7.20. A ocorrência desse tipo de sujeito é muito semelhante nos três *corpora*: 460 na Amostra 2; 457 no DG; e 464 no JM. Isso reafirma os achados de Finatto *et al.* (2011) sobre as elipses em jornais populares. Segundo esse estudo por nós já mencionado nas considerações finais do Estudo-Piloto 1 e também nas análises desse tipo de sujeito no Estudo-Piloto 3, a presença de elipses não é um fator decisivo na avaliação da complexidade textual.

Cumpramos também repetir que os sujeitos elípticos na Amostra 2, tal qual na Amostra 1, encontram-se em primeiro lugar no *ranking*. Esse dado, além de reafirmar a representatividade da amostra, nos permite inferir que as sentenças nesse quesito, também sejam mais complexas que os textos dos jornais. Como já mencionamos nas análises da Amostra 1, mesmo não sendo fator decisivo para a maior ou menor complexidade textual, esses sujeitos exigem mais da memória de trabalho do leitor.

b) **Sujeitos Pré-Verbais**

Entre os tipos de CFSs com maior ocorrência nas amostras, verificamos que, nos jornais, os sujeitos pré-verbais aparecem em primeiro lugar no *ranking*. Considerando que sujeitos pré-verbais representam a ordem canônica da língua – sujeito, verbo e complementos -, os textos dos jornais DG e JM, nesse quesito, apresentam-se adequados ao perfil de leitor a que se destinam. Assim se dá, porque a ordem canônica da língua, por ser mais natural entre os falantes dessa língua, é mais fácil de ser compreendida.

Na Amostra 2, os sujeitos pré-verbais ocupam o segundo lugar no *ranking*. Mesmo estando nessa colocação, o número de ocorrências desse tipo de sujeito é praticamente a

metade do que ocorre nos jornais. Em percentuais relacionados à totalidade de sujeitos de cada *corpora*, os pré-verbais apresentam-se da seguinte forma: 32,03% na Amostra 2; 45,31% no DG; e 47,18% no JM. Ou seja, as Sentenças apresentam em torno de 15% a menos de sujeitos pré-verbais que os jornais.

c) **Sujeitos Pós-Verbais**

Em relação aos sujeitos pós-verbais, o processo se inverte em maiores proporções. Ou seja, a Amostra 2 apresenta praticamente o dobro de ocorrências de sujeitos pós-verbais que os jornais: 200 na Amostra 2, 97 no DG; e 92 no JM. Em termos percentuais relacionados à totalidade de sujeitos de cada *corpora*, temos os seguintes dados: 16,9% na Amostra 2; 6,8% no DG; e 6,4% no JM.

A diferença desses percentuais é maior que o dobro na Amostra 2 em relação aos jornais. Esse é um dado bastante significativo, porque, como já referimos, a alteração da ordem de colocação do sujeito é um dificultador para o processamento da leitura. Em relação à posição do sujeito, as Sentenças, então, apresentam-se mais complexas que os textos dos jornais.

Cumprindo, porém, observar, que uma das possíveis razões para a maior ocorrência de sujeitos pós-verbais nas sentenças esteja no fato de, nesses textos, serem frequentes frases como as dos exemplos abaixo:

- (1) Vieram os autos conclusos.
- (2) Foi dada vista à parte-autora.
- (3) Intimem-se as partes.

Os dois primeiros narram um evento ocorrido no processo, e o último expressa um comando que ocorre na parte final da sentença.

O primeiro exemplo, com um verbo intransitivo, e o terceiro, na voz passiva sintética, estão entre os casos de sujeitos pospostos mencionados por gramáticos como Bechara e Cunha e Cintra.

O Quadro 7.3 traz exemplos de sujeitos pós-verbais das sentenças da Amostra 2⁹⁷ com a quantidade de ocorrências (os sujeitos estão sublinhados, e os seus verbos em negrito).

⁹⁷ Não tivemos tempo hábil para a análise na Amostra 1.

Quadro 7.3 – Exemplos de CFSs Pós-Verbais das Sentenças da Amostra 2

Estão presentes <u>os pressupostos caracterizadores da responsabilidade civil e, consequentemente, do direito à reparação [...]</u> .
O verbo em destaque é de ligação. (43 ocorrências)
Inexiste <u>condenação em custas e honorários no primeiro grau.</u>
O verbo em destaque é um verbo intransitivo. (41 ocorrências)
Encerrada <u>a instrução, vieram os autos para parecer.</u>
O primeiro verbo em destaque no exemplo acima está no particípio e introduz uma oração subordinada adverbial. (35 ocorrências)
O segundo é outro exemplo de verbo intransitivo.
Foi deferido (<i>sic</i>) <u>a gratuidade de justiça</u> ao requerente e indeferido (<i>sic</i>) <u>a tutela de urgência</u> (fls. 41).
Os verbos em destaque estão na voz passiva analítica. (32 ocorrências)
Na data prevista na audiência de instrução, considerem-se <u>todos intimados[...]</u> .
O verbo em destaque está na voz passiva sintética. (18 ocorrências)
Assim, não tendo havido qualquer prova que colida com o constante do boletim de acidente de trânsito, merece <u>acolhimento a conclusão nele havida.</u>
O verbo em destaque é transitivo. (13 ocorrências)
Ainda, declara <u>a requerente</u> que realizou a referida compra com o crédito fornecido pela segunda requerida [...]
O verbo em destaque é de elocução e ocorrem principalmente no relatório da sentença. (10 ocorrências)

As posições ocupadas pelos verbos intransitivos (2º lugar) e verbos transitivos (6º ou penúltimo lugar) no Quadro 7.3 vão ao encontro das conclusões de Pilati (2002, p. 25): “as orações declarativas do PB com ordem VS apresentam dois padrões de comportamento: (i) inversão menos restrita, para orações com verbos inacusativos⁹⁸, (ii) inversão mais restrita se o verbo for inergativo ou transitivo”.

⁹⁸ Não entraremos aqui na diferenciação entre verbos inacusativos e inergativos por não ser o foco da nossa pesquisa. Pela Gramática Tradicional, são considerados intransitivos.

Essa autora menciona que algumas pesquisas sobre o tema afirmam que, entre os verbos intransitivos, verbos com sentido apresentativo como *acontecer*, *aparecer*, *vir* são os que mais ocorrem com sujeitos pospostos. Na Amostra 2, o verbo *vir* ocorreu 5 vezes, em todas elas com sujeito posposto, e o verbo *advir* (derivado de *vir*) ocorreu 1 vez, também com sujeito posposto.

d) Sujeitos Oracionais

Destacamos este tipo de sujeito em razão de sua apresentação em forma de oração. Em princípio, a oração pode representar um constituinte mais extenso em número de palavras e também mais complexo porque em sua estrutura interna também abriga outros constituintes frasais.

A Tabela 7.20 mostra que o número de ocorrência desse tipo de sujeito na Amostra 2 em relação à totalidade de seus sujeitos não é muito significativo: 45 para 1177. Entretanto, em relação às ocorrências nos jornais, o número de sujeitos oracionais das sentenças supera em mais de 30%. Esse índice pode também ser um indicativo da maior complexidade textual das sentenças em relação aos jornais nesse quesito.

Se somarmos o número de sujeitos oracionais com o número de sujeitos elípticos e pós-verbais das sentenças, temos um total de 704, que, em termos percentuais em relação à totalidade de sujeitos da Amostra 2, representa 59,81% de sujeitos que demandam maior sobrecarga de trabalho na leitura. Essa soma nos jornais gera os seguintes dados: 40,25% no DG; e 39,55% no JM. Tais percentuais novamente indicam a maior complexidade textual das sentenças em relação aos jornais nesse quesito.

e) Sujeitos Pronominais

Quanto aos sujeitos pronominais, inicialmente destacamos que os jornais apresentam mais que o dobro das ocorrências da Amostra 2. Os sujeitos pronominais também demandam uma maior sobrecarga de trabalho no processamento da leitura, por necessitarem de um referente para serem interpretados. Assim, em princípio os textos dos jornais populares ficaram mais complexos que os das sentenças nesse quesito.

Após essa classificação inicial, adotamos os mesmos dois procedimentos realizados na Amostra 1. Vamos a eles.

Procedimento 1:

Distribuímos os sujeitos pré-verbais, pós-verbais e oracionais por número de palavras para verificar a extensão das dependências sintáticas: o número de palavras entre

a raiz sintática (no caso, o SV) e o dependente (no caso, o constituinte com função de sujeito). A Tabela 7.21 apresenta esses dados.

Tabela 7.21 – CFSs por Número de Palavras na Amostra 2, DG e JM

TIPOS	CORPORA	NÚMERO DE OCORRÊNCIAS POR QUANTIDADE DE PALAVRAS						TOTAL
		01-05	06- 10	11-20	21-30	31-40	41+	
PRÉ-VERBAL	AMOSTRA 2	326	33	17	1	0	0	377
	DG	510	83	42	3	0	0	638
	JM	533	112	28	7	0	0	680
PÓS-VERBAL	AMOSTRA 2	108	48	39	2	3	0	200
	DG	69	21	7	0	0	0	97
	JM	60	21	9	1	0	1	92
ORACIONAL	AMOSTRA 2	2	11	18	9	1	3	44
	DG	6	2	3	1	0	0	12
	JM	4	7	3	1	1	0	16
TOTAL		1618	338	166	25	5	4	2156

Pelo total exibido na última linha da Tabela 7.21, sujeitos com até 5 palavras representam 75,18% do total de sujeitos pré-verbais, pós-verbais e oracionais. E, à medida que aumenta o número de palavras, o número de ocorrências diminui em todas as amostras, com exceção de duas situações encontradas na Amostra 1. Nessa amostra, ocorre um sujeito a mais entre os sujeitos pós-verbais na faixa de 31 a 40 palavras do que na faixa de 21 a 30; e duas ocorrências a mais nos sujeitos oracionais, na faixa de 41 palavras ou mais em relação à faixa de 31 a 40 palavras.

A seguir apresentamos algumas observações sobre essas ocorrências de CFSs.

a) Sujeitos Pré-Verbais:

Entre os sujeitos pré-verbais, ainda que nas sentenças as ocorrências sejam menores do que nos jornais, algumas situações chamam a atenção. São elas:

- a. Na faixa de 0 a 5 palavras. Na Amostra 2, o percentual de ocorrências em relação ao total exibido na última coluna e mesma linha é de 86,47%, contra 79,93% no DG e 78,38% no JM.
- b. O número de ocorrências de sujeitos na faixa de 6 a 10 palavras. Na Amostra 2, temos um percentual de 8,75% de ocorrências contra 13% no DG e 16,47% no JM.

- c. O número de ocorrências de sujeitos na faixa de 21 a 30 palavras. Em termos percentuais, 7 ocorrências no JM representam 1,02% contra 1 ocorrência, que representa 0,26% desse tipo de sujeito nas Sentenças.

Tais dados indicam que as sentenças apresentam maior ocorrência de CFSs pré-verbais mais simples, com menos palavras, que os jornais.

No Quadro 7.4, colocamos alguns exemplos desses CFSs (parte sublinhada).

Quadro 7.4 – Exemplos de CFSs Pré-Verbais das Amostras

AMOSTRA	EXEMPLO
AMOSTRA 2	a) <u>Proposição nº 3 – O cadastramento indevido em órgãos de restrição ao crédito é causa por si só de indenização por danos morais, quando se tratar de única inscrição e, de forma excepcional, quando houver outras inscrições regulares.</u> (09 palavras)
	b) <u>Afigura-se despicienda a discussão a respeito da aplicação da Lei 9.656/98 à hipótese, tendo em vista que o fundamento utilizado pelo acórdão recorrido, referente à análise das cláusulas contratuais em conformidade com o diploma consumerista, (sic) é suficiente, por si só para mantê-lo.</u> (18 palavras)
DG	c) <u>As inscrições na Copa Ênio Costamilan e no Brasileiro Sub-23 também não puderem ser feitas a tempo.</u> (10 palavras)
	d) <u>Uma grávida de 38 anos, o filho dela, de 13 anos, e o irmão da mulher, de 39 anos, ficaram feridos.</u> (19 palavras)
JM	e) <u>A pasta respondeu que a discussão acerca dos reflexos da pandemia no transporte público vem sendo mantida entre a Frente Nacional de Prefeitos (FNP) e o Governo Federal, visando buscar possíveis soluções.</u> (10 palavras)
	f) <u>Após dois anos sendo publicada apenas na versão digital, a revista Exclusiva, com 32 anos de história, pioneira em cobrir o Carnaval de Salvador, micaretas, eventos juninos e outros, volta nesta segunda-feira (30) com versão impressa e com formato de variedades.</u> (20 palavras)

No Quadro 7.4, chama a atenção a extensão dos trechos exemplificados no DG: bem menores que os das outras amostras. Este comportamento confere com os dados

relativos ao número de palavras por frase em cada uma das amostras (Tabela 7.19), em que o DG apresentou um índice bem menor que o JM e as Sentenças. Além disso, os dois exemplos do DG trazem o CFS como primeiro elemento da frase, o que é um facilitador para a leitura. Nos dois exemplos do JM, como se vê, os sujeitos se encontram na parte interna do trecho, em uma das orações que constitui a frase em questão.

b) Sujeitos Pós-Verbais:

Entre os sujeitos pós-verbais, como já vimos, a maior ocorrência se dá nos textos das Sentenças, com praticamente o dobro de ocorrências do que nos jornais. De todo modo, aqui também alguns dados chamam a atenção.

a) Na faixa de 0 a 5 palavras. As sentenças, com 108 ocorrências, apresentam um percentual de 54% do total de sujeitos nessa faixa. Os jornais, por sua vez, em termos percentuais, apresentam os seguintes índices: DG, 71%, com 69 ocorrências; e JM, 65%, com 60 ocorrências.

b) Na faixa de 6 a 10 palavras. Os percentuais em relação à totalidade de sujeitos em todas as amostras ficaram muito próximos: Amostra 2, com 24%; DG, com 21,54%; e JM, com 22,83%.

c) Na faixa de 11 a 20 palavras. O percentual de ocorrência nas sentenças ficou mais que o dobro que o dos jornais: 19,5% contra 7,21% no DG e 9,7% no JM.

No Quadro 7.5, colocamos alguns exemplos desses CFSs (parte sublinhada).

Quadro 7.5 – Exemplos de CFSs Pós-Verbais das Amostras

AMOSTRA	EXEMPLO
AMOSTRA 2	a) Configurada <u>a falha na prestação de serviços por parte da demandada</u> , restam reconhecidos como verdadeiros <u>os fatos alegados pela autora</u> , o que culmina na fixação de danos morais na soma de R\$ 5.000,00. (10 e 5 palavras)
	b) Notadamente diante da jurisprudência deste Tribunal, que já se consolidou no sentido de que é abusiva <u>a cláusula restritiva de direito que exclui do plano de saúde o custeio de prótese em procedimento cirúrgico coberto pelo plano e necessária ao pleno restabelecimento da saúde do segurado</u> , sendo indiferente, para tanto, se referido material é ou não importado. (30 palavras)
DG	c) Em três meses, queremos ter esse quadro – adianta <u>o gerente de pacientes externos do Conceição, João Albino Potrich</u> .

	(10 palavras)
	d) Foi adiado para 8 de julho deste ano <u>o julgamento dos réus Eduardo Carvalho Lyra, o Dudu, 26 anos, e Fernando Silveira da Rosa, o Rê, 26 anos.</u>
	(20 palavras)
JM	e) Antes de abril deste ano, quando começou <u>o maior rigor do isolamento domiciliar causado pela pandemia</u> , a Defensoria atendia cerca de 200 mulheres por mês.
	(9 palavras)
	f) Nesse período, para quem tem problema respiratório, se aconselha <u>melhor ventilação do ambiente e não ter objetos que concentrem fungos e ácaros como tapetes, carpetes, evitar o uso de ventiladores direcionados à face.</u>
	(24 palavras)

Com esses dados, é possível verificar que os textos das Sentenças confirmam a sua maior complexidade em relação aos textos dos jornais populares. Além de apresentarem mais ocorrências desse tipo de sujeito, a extensão desses constituintes frasais em relação aos jornais é também maior. A maioria dos sujeitos pós-verbais nos jornais fica na faixa de 0 a 5 palavras com um percentual bem maior do que nas sentenças. E na faixa de 11 a 20 palavras, a diferença percentual entre as ocorrências das Sentenças e a dos jornais é bem significativa: as sentenças apresentam uma ocorrência de 19,5% contra 7% para o DG e 9,7% para o JM.

Chama a atenção novamente aqui a extensão dos exemplos do DG, trechos bem menores que os das outras amostras.

c) Sujeitos Oracionais

Como já observamos, a ocorrência de sujeitos oracionais nas Sentenças é mais que três vezes superior à que se dá nos jornais. Além disso, observa-se que a maior incidência desse tipo de constituinte frasal na Amostra 2 se dá na faixa de 11 a 20 palavras. Chama também a atenção a presença de 9 sujeitos oracionais na faixa de 21 a 30 palavras e 3 com mais de 41 palavras.

No Quadro 7.6, a seguir, apresentamos alguns exemplos de sujeitos oracionais (parte sublinhada) com o respectivo número de palavras em nossas amostras.

Quadro 7.6 – Exemplos de CFSs Oracionais das Amostras

AMOSTRA	EXEMPLO
AMOSTRA 2	<p>a) Todavia, destaca-se <u>que a relação de consumo que opera a inversão do ônus da prova, nos termos do artigo 6º, inciso VIII, do CDC, não importa em desonerar a parte autora da comprovação mínima de suas alegações e dos fatos constitutivos do seu direito, consoante artigo 373, inciso I, do CPC.</u></p> <p style="text-align: right;">(42 palavras)</p>
DG	<p>b) É claro <u>que a forte concorrência da dupla campeã da Libertadores – Bolívar e Índio – é garantia de que não vai sobrar espaço.</u></p> <p style="text-align: right;">(20 palavras)</p>
JM	<p>c) Recomenda-se <u>que crianças de até 11 anos realizem pelo menos 60 minutos de atividade física por dia, com duas horas ou menos de uso de mídia de tela e durmam de nove a 11 horas.</u></p> <p style="text-align: right;">(34 palavras)</p>

Novamente aqui a extensão do exemplo extraído do DG é bem menor que a das outras amostras.

Considerando encerrada a análise por número de palavras, passamos ao Procedimento 2.

Procedimento 2:

Nesta etapa, tal qual foi feito com a Amostra 1, verificamos a forma como se apresentavam os sujeitos elípticos, os sujeitos oracionais e os sujeitos pronominais. Apresentamos nossas observações por tipo de sujeito.

a) Sujeitos Elípticos

As formas verbais relativas aos sujeitos elípticos de nossos *corpora* apresentam-se conforme consta na Tabela 7.22, a seguir.

Tabela 7.22 – Formas Verbais Referentes aos CFSs Elípticos

SUJEITOS ELÍPTICOS		
FORMAS VERBAIS	CORPORA	NÚMERO DE OCORRÊNCIAS
Verbos na 1ª pessoa	AMOSTRA 2	79
	DG	104
	JM	111
Verbos na 3ª pessoa	AMOSTRA 2	230
	DG	185
	JM	161
Verbos no Infinitivo	AMOSTRA 2	54
	DG	100
	JM	112
Verbos no Gerúndio	AMOSTRA 2	55
	DG	21
	JM	35
Verbos no Particípio	AMOSTRA 2	39
	DG	22
	JM	25
Outros (Imperativo 2ª pessoa)	AMOSTRA 2	0
	DG	26
	JM	13

Entre os sujeitos elípticos representados por verbos flexionados, os jornais apresentam uma maior incidência de verbos na 1ª pessoa e as Sentenças de verbos na 3ª pessoa. Entre os 230 CFSs elípticos representados por verbos na 3ª pessoa nas sentenças, 211 estão flexionados na 3ª pessoa do singular. Como vimos, verbos na 3ª pessoa do singular podem apresentar maior número de candidatos a referentes do que os da 3ª do plural. Isso demanda mais trabalho do leitor e, portanto, torna o texto mais complexo. Diante disso, nesse quesito os jornais apresentam-se menos complexos que as Sentenças.

Além disso, cumpre referir que, na Amostra 1, os sujeitos elípticos representados por verbos flexionados na 3ª pessoa foram também os de maior ocorrência. Esse dado confirma mais uma vez a representatividade da Amostra 2.

Em relação aos sujeitos elípticos representados por verbos nominais, a amostra com maior ocorrência no total foi o JM, com 172, contra o DG, com 143, e as Sentenças com 148. As formas nominais também oferecem um trabalho a mais para o processamento do seu referente. Nesse quesito, então, o JM parece ser o mais complexo.

O Quadro 7.7 apresenta exemplos de sujeitos elípticos com problemas na identificação do referente.

Quadro 7.7 – Exemplos de CFSs Elípticos das Amostras

AMOSTRA	EXEMPLO
AMOSTRA 2	a) A modificação pretendida pela autora altera toda a base contratual, sendo necessária análise de percentuais cobrados, com eventual restituição de quantias pagas, necessitando de cálculo para o reembolso da diferença, se por acaso, constatar-se a abusividade.
	Neste trecho, é difícil encontrar o sujeito do gerúndio “necessitando”. A ideia do trecho parece ser dizer que é necessário analisar os percentuais cobrados, com eventual restituição de quantias pagas, e calcular o reembolso da diferença, se algum abuso for constatado.
DG	b) A CPI do Detran prepara uma sessão para divulgar o conteúdo de escutas telefônicas feitas pela Polícia Federal.
	Neste trecho, o infinitivo divulgar apresenta dois candidatos a sujeito: a CPI do Detran , ou uma sessão .
JM	c) A proprietária do andar térreo me encaminhou um bilhete dizendo que um técnico irá avaliar a caixa d'água dela, que se encontra em minha laje (ao lado da caixa d'água do primeiro piso, onde ficam as duas).
	Neste caso, o gerúndio dizendo pode ter dois candidatos a sujeito: a proprietária do andar térreo e bilhete .

b) Sujeitos Oracionais

A Tabela 7.23 mostra como se apresentam os sujeitos oracionais: se na forma de orações desenvolvidas, com a presença da conjunção, ou se reduzidas, em uma das formas nominais do verbo.

Tabela 7.23 – Tipos de CFSs Oracionais

AMOSTRA	NÚMERO DE OCORRÊNCIAS	
	Oração desenvolvida (com a conjunção)	Oração reduzida
AMOSTRA 2	29	16
DG	3	9
JM	6	7
TOTAL	38	32

Em relação aos sujeitos oracionais, a Amostra 2 novamente reflete o ocorrido na Amostra 1, com maior índice de sujeitos oracionais representados por orações desenvolvidas do que orações reduzidas.

Comparando com os textos do DG e do JM, já referimos anteriormente que as Sentenças apresentaram três vezes mais sujeitos representados por orações do que os jornais. Isso também é indício de maior complexidade leitora, uma vez que as orações tendem a ser mais extensas que os sintagmas nominais. Apesar disso, as orações desenvolvidas ocorrem com maior frequência em todas as amostras, o que é um fator positivo. Assim pode ser considerado, porque as reduzidas, por estarem em uma das formas nominais do verbo, demandam maior trabalho na busca do referente de cada uma dessas formas verbais.

O Quadro 7.8 apresenta exemplos de sujeitos oracionais nas formas desenvolvidas e reduzidas.

Quadro 7.8 – Exemplos de Tipos de CFSs Oracionais das Amostras

AMOSTRA	EXEMPLO
AMOSTRA 1	a) Nosso ordenamento jurídico engloba a teoria finalista mitigada para a interpretação da relação existente entre adquirente do produto/serviço e fornecedor/fabricante, devendo ser analisado cada caso em concreto para se verificar <u>se há a existência ou não de relação de consumo</u> . (oração desenvolvida)
	b) Desta forma, cabia à demandada fazer prova da origem do <u>crédito questionado</u> , ônus do qual não se desincumbiu, não trazendo aos autos qualquer prova da efetiva utilização do serviço. (oração reduzida de infinitivo)
DG	c) Certo é que Roth vive situação instável. (oração desenvolvida)
	d) Para ser bombeiro, é preciso fazer concurso público e <u>passar por um curso</u> . (oração reduzida de infinitivo)
JM	e) É aconselhável que os pais procurem um pediatra, médico otorrinolaringologista ou fonoaudiólogo <u>quando houver alguma suspeita de problemas auditivos na criança</u> . (oração desenvolvida)
	f) É importante higienizar a região do braço até o cotovelo. (oração reduzida de infinitivo)

c) Sujeitos Pronominais

A Tabela 7.24 apresenta as ocorrências dos CFSs por tipos de pronomes.

Tabela 7.24 – Tipos de CFSs Pronominais

TIPOS DE PRONOMES	NÚMERO DE OCORRÊNCIAS POR <i>CORPORA</i>		
	AMOSTRA 2	DG	JM
Relativos (que, quem, qual)	79	119	120
Pessoal	0	71	54
Demonstrativo	8	4	8
Indefinido	9	8	9
TOTAL	96	202	191

No caso dos CFSs pronominais na Amostra 2, repete-se o que ocorreu na Amostra 1: os pronomes relativos ficaram em primeiro lugar no *ranking* de ocorrências. Mais um dado a confirmar a representatividade da Amostra 2.

Os pronomes relativos representam a maioria dos sujeitos pronominais em todos os *corpora*. Nas Sentenças, representam 82,29%; no DG, 58,91%, e, no JM, 62,83%. Assim, ainda que o número bruto de sujeitos pronominais seja maior nos jornais do que na Amostra 2, nas Sentenças a proporção de sujeitos representados por pronomes relativos e, conseqüentemente, de orações encaixadas, é maior do que nos jornais. Quanto a isso, então, as sentenças podem ser consideradas mais complexas que os jornais.

Cumpra ainda referir que Vilela (1988), na sua dissertação de mestrado sobre o sujeito anafórico elíptico, encontrou muito poucos sujeitos pronominais em seu *corpus* (textos de jornais). Ele denomina esses elementos de Sujeitos Anafóricos Pronominais (SAPro) e explica que “o uso reiterado de pronomes-sujeito é visto como ‘pobreza de estilo’. Portanto, apesar de toda a ambigüidade potencial das formas verbais de terceira pessoa, usam-se muito pouco, em língua escrita, os pronomes-sujeito”. (VILELA, 1988, p. 31).

Tendo em conta essa afirmação de Vilela, o DG e o JM parecem não estar muito preocupados com a pobreza de estilo, e os autores das Sentenças, sim. Porém, no caso da Amostra 2, a ausência de pronomes pessoais retos na função de sujeito pode ser explicada pelo fato de que os atores dos textos das Sentenças são designados pelas funções que exercem no contexto do processo relativo a essas Sentenças. Por exemplo, normalmente quem propõe a ação é designado como “parte autora”, “autor/autora”, “demandante”, etc. E aquele contra quem se propõe a ação é designado como “parte ré”, “réu/ré”, “requerido/requerida”, “demandado/demanda”. Assim, a referência que se faz a um desses atores é sempre por essas designações e não por um pronome pessoal, como

“ele(s)” ou “ela(s)”. Quanto ao pronome “eu”, que representaria a voz do autor da sentença, aparece sempre subentendido e foi relacionado entre os sujeitos elípticos.

No Quadro 7.9, apresentamos alguns exemplos desse tipo de sujeitos.

Quadro 7.9 – Exemplos de CFSs Pronominais das Amostras

AMOSTRA	EXEMPLO
AMOSTRA 2	a) Incontroversa, nos autos, a relação contratual entre as partes, que acarretou as dívidas de folhas 19, com vencimento entre 26/5/2014 a 29/7/2014.
	b) Conta que, em razão da dor da perda, optou por não realizar a festa e que, ao comunicar o fato para o réu, este não concordou com a devolução do valor pago pelos serviços.
DG	c) Mas foi o atual camisa 1 gremista, Victor, que se destacou no sábado. Ele evitou a derrota na Ressacada com três grandes defesas em menos de dez minutos.
JM	d) O humorista baiano João Pimenta foi contratado pelo Porta dos Fundos. No seu Instagram, ele deu a notícia.
	e) Gostaria de chamar atenção para a escuridão que , todo dia, assusta os moradores que residem nas proximidades da localidade da Praça Nossa Senhora Aparecida, no bairro do Imbuí.

7.2.2.2 Constituintes Frasais com Função de Complementos Verbais (CFCVs)

Para a extração desses CFCVs da Amostra 2, selecionamos os que já haviam sido previamente analisados para a descrição da Amostra 1. Em relação aos textos dos jornais, lemos novamente cada uma das orações de nossos *corpora*. Por fim, para este Estudo-Piloto ficamos com um total de 3.630 CFCVs. Para a análise desses constituintes, adotamos alguns procedimentos.

Procedimento 1:

Esses CFCVs foram divididos, num primeiro momento, tal qual foi feito com a Amostra 1 de sentenças, em três grupos.

- a) complementos verbais preposicionados (CFCVPs);
- b) complementos verbais não preposicionados (CFCVNPs); e
- c) complementos verbais representados por pronomes (CFCVPRs).

Os CFCVPs estão relacionados ao verbo por meio de uma preposição, e os CFCVNPs complementam o verbo sem a intermediação de preposição. Os CFCVPRs, foram colocados à parte, mesmo que alguns possam vir acompanhados de preposição.

Entre os CFCVPs, encontram-se complementos qualificados pela Gramática Tradicional como objeto indireto, predicativo, complemento circunstancial e agente da passiva. Entre os CFCVNPs, encontram-se complementos verbais classificados como objeto direto e predicativo. E entre os CFCVPR também encontramos essas categorias de complementos verbais.

Na Tabela 7.25, apresentamos a distribuição quantitativa desses CFCVs em nossa Amostra 1

Tabela 7.25 – Constituintes Frasais com Função de Complemento Verbal (CFCVs)

TIPOS	OCORRÊNCIAS			TOTAL
	Amostra 2	DG	JM	
PREPOSICIONADO - CFCVP	327	334	298	959
NÃO PREPOSICIONADO - CFCVNP	758	834	855	2447
PRONOMINAIS - CFCVPR	62	89	73	224
TOTAL	1147	1257	1226	3630

Pelos dados da Tabela 7.25, verificamos mais que o dobro de CFCVNPs em relação aos CFCVPs. Isso também aconteceu com a Amostra 1 das sentenças. Quanto aos CFCVPRs, as três amostras apresentaram percentuais semelhantes de ocorrência: Amostra 2, com 5,44%; DG, com 7,03%; e JM, com 5,95%. O percentual de ocorrência desses constituintes na Amostra 1 de sentenças foi igualmente semelhante: 6,28% de CFCVPRs.

Após a distribuição dos CFCVs da Tabela 7.25, subdividimos, tal qual fizemos com a Amostra 1, esses complementos de acordo com sua estrutura: em sintagmas nominais e sintagmas oracionais. Os CFCVPRs foram subdivididos conforme o tipo de pronome.

A Tabela 7.26 traz os dados dessa subdivisão.

Tabela 7.26 – CVFCVs e seus Subtipos Presentes nas Amostras 2, DG e JM

TIPO	SUBTIPO	NÚMERO DE OCORRÊNCIAS POR <i>CORPORA</i>			
		Amostra 2	DG	JM	TOTAL
CFCVP	SINTAGMAS NOMINAIS	313	321	288	922
	SINTAGMAS ORACIONAIS	14	13	10	37
CFCVNP	SINTAGMAS NOMINAIS	628	756	730	2114
	SINTAGMAS ORACIONAIS	130	78	125	333
CFCVPR	relativo	16	23	18	57
	oblíquo (lhe, me, o, a, se)	46	62	49	157
	reto	0	3	4	7
	indefinido	0	1	2	3
TOTAL		1147	1257	1226	3630

Entre os subtipos de CFCVs apresentados na Tabela 7.26, verificamos que 3034 são representados por sintagmas nominais contra 370 ocorrências de sintagmas oracionais. Isso, em termos percentuais em relação à totalidade de CFCVs dos *corpora analisados*, gera 83,63% de CFCVs na forma de sintagmas nominais e 10,19% de CFCVs na forma de sintagmas oracionais. Esses dados ficam muito próximos do que ocorreu com a Amostra 1. Nessa amostra, o total de 8.847 CFCVs na forma de sintagmas nominais representam 81,51% do total de CFCVs nela encontrados. E o total de 1259 CFCVs oracionais representa 11,61%.

A Amostra 2, por sua vez, com 941 ocorrências de CFCVs sintagmas nominais num total de 1147 CFCVs, apresenta um percentual de 82,04%. E as 145 ocorrências de CFCVs oracionais representam um percentual de 12,64%. Tais percentuais são muito próximos daqueles que ocorreram na Amostra 1, o que é mais um dado a confirmar a representatividade das amostras analisadas.

Quanto aos CFCVPRs, chama a atenção não haver nenhuma ocorrência de pronome reto nessa condição na Amostra 2. Na Amostra 1, ocorreu 16 vezes. Os pronomes indefinidos não tiveram ocorrência nas duas amostras.

Em todos os *corpora* observa-se a maior ocorrência de CFCVPRs representados por pronomes oblíquos, seguidos dos representados por pronomes relativos. Nesse último caso, aparentemente a amostra de Sentenças ficou com um menor número de ocorrências. Porém, em termos percentuais em relação à totalidade de cada uma das amostras, as

Sentenças e o JM ficaram com 1,4%. O DG é que teve um percentual um pouco maior: 1,83%. A Amostra 1 ficou com um percentual de 1,7% em relação ao total de CFCVs.

A seguir, no Quadro 7.10, apresentamos exemplos de CFCVs segundo a classificação da Tabela 7.26 das nossas amostras.

Quadro 7.10 – Exemplos de CFCVs das Amostras

AMOSTRA	EXEMPLO
AMOSTRA 2	a) Com efeito, a requerida Serasa (a) <u>se</u> desincumbiu (b) <u>de comprovar o envio de notificação prévia ao consumidor inscrito nos órgãos restritivos de crédito</u> , conforme os documentos juntados que acompanham (c) <u>a defesa</u> , satisfazendo (d) <u>a exigência do disposto no art. 43, §2, do CDC, e na Súmula 359 do STJ</u> , restando (e) <u>afastada</u> qualquer obrigação na indenização pleiteada, referente a referida ré.
	As partes sublinhadas no trecho acima são exemplos de: (a) CFCV pronominal (b) CFCVP oracional (c) CFCVNP sintagma nominal (d) CFCVNP sintagma nominal (e) CFCVNP sintagma nominal
DG	b) Quando ocorre uma variação na temperatura, as doenças pulmonares pioram e (a) <u>se</u> somam (b) <u>a todas as outras doenças que causam a superlotação</u> – explica o chefe das emergências da Santa Casa, Leonardo Fernandez.
	As partes sublinhadas no trecho acima são exemplos de: (a) CFCV pronominal; e (b) CFCVP sintagma nominal No CFCVP da letra b, temos um sintagma nominal qualificado por uma oração relativa – que causam a superlotação. Esta oração relativa foi computada na totalidade do sintagma nominal.
JM	c) Daí esse dinheiro entra (a) <u>no caixa das empresas que passam a honrar (b) seus compromissos com a folha de pagamento e com a compra de combustível para manter (c) o serviço</u> ”, explicou.
	As partes sublinhadas no trecho acima são exemplos de: (a) CFCVP sintagma nominal (b) CFCVNP sintagma nominal (c) CFCVNP sintagma nominal No CFCVP da letra (a), ocorre novamente temos um sintagma nominal qualificado por uma oração relativa – que passam a [...] combustível. A oração “para manter o serviço”, por estar subordinada à “combustível”, foi incluída na totalidade desse CFCVNP.

Após essa subdivisão por tipos, adotamos um segundo procedimento, que a seguir passamos a explicar.

Procedimento 2:

Nesse procedimento, distribuímos os CFCVs de nossas amostras por número de palavras. Tal qual fizemos com a Amostra 1, a ideia é verificar a extensão das dependências sintáticas, ou seja, o número de palavras entre a raiz sintática (no caso, o SV) e o dependente (no caso, o constituinte frasal com função de complemento verbal).

A Tabela 7.27 apresenta os dados relativos aos CFCVs constituídos por sintagmas nominais, e a Tabela 7.29 apresenta os dados relativos aos CFCVs constituídos por sintagmas oracionais. Os CFCVPRs não entraram nessa distribuição.

Tabela 7.27 – CFCVs Sintagmas Nominais por número de palavras

TIPO DE CFCV	CORPORA	OCORRÊNCIA POR NÚMERO DE PALAVRAS						
		1-5	6-10	11-20	21-30	31-40	41+	TOTAL
CFCVNP	AMOSTRA 2	200	67	31	9	3	3	313
	DG	235	48	32	5	1	0	321
	JM	195	57	31	3	2	0	288
CFCVNP	AMOSTRA 2	417	124	73	7	4	3	628
	DG	537	141	60	16	2	0	756
	JM	488	146	74	15	4	3	730
TOTAL		2072	583	301	55	16	9	3036

Na última linha da Tabela 7.27, é possível verificar que mais da metade dos CFCVs das três amostras analisadas encontra-se na faixa de 1 a 5 palavras, diminuindo a ocorrência à medida que o número de palavras aumenta. Isso também aconteceu com a Amostra 1.

A fim de melhor compararmos as amostras, colocamos abaixo a Tabela 7.28 com as frequências da Tabela 7.27 em percentuais.

Tabela 7.28 – Percentuais de CFCVs Sintagmas Nominais por Número de Palavras

TIPO DE CFCV	CORPORA	PERCENTUAIS DE OCORRÊNCIA POR NÚMERO DE PALAVRAS					
		1-5	6-10	11-20	21-30	31-40	41+
CFCVP	AMOSTRA 2	17,4%	5,8%	2,7%	0,8%	0,26%	0,26%
	DG	18,7%	3,8%	2,5%	0,4%	0,07%	0%
	JM	15,90%	4,64%	3,5%	0,2%	0,16%	0%
CFCVNP	AMOSTRA 2	36,3%	10,8%	6,4%	0,6%	0,35%	0,26%
	DG	42,7%	11,2%	4,7%	1,3%	0,16%	0%
	JM	39,80%	11,09%	6%	1,2%	0,32%	0,24%

Os percentuais colocados na Tabela 7.28 são relativos à totalidade de CFCVs presentes em cada amostra. No caso da Amostra 2, por exemplo, os percentuais foram calculados sobre o total de 1.147 CFCVs nela encontrados, conforme Tabelas 7.25 e 7.26. De modo geral, os percentuais das três amostras são muito próximos em todas as faixas de ocorrências de palavras. Alguns dados, porém, chamam a atenção:

- Entre os CFCVNPs, na faixa de 21 a 30 palavras, as Sentenças ficaram com um percentual menor que o dos jornais, mas o número de ocorrências é igual ao que se deu no JM: 3.
- Entre os CFCVPs, na faixa de 11 a 20 palavras, o JM apresenta um percentual superior ao das Sentenças.
- Entre os CFCVPs e os CFCVNPs, na faixa de 41 palavras ou mais, o DG não apresenta nenhuma ocorrência.

No Quadro 7.11 selecionamos, de cada uma das amostras, um exemplo de CFCV com mais de 30 palavras.

Quadro 7.11 – Exemplos de CFCVs das Amostras com mais de 30 Palavras

AMOSTRA	EXEMPLO
AMOSTRA 2	a) ANTE O EXPOSTO, opino <u>pela PARCIAL PROCEDÊNCIA dos pedidos para condenar a ré à indenização por danos morais de R\$2.000,00 (dois mil reais), valor a ser corrigido pelo IGP-M, a contar da data da homologação deste parecer, e acrescido de juros de 1% ao mês, a partir da citação.</u>
	O CFCV sublinhado no exemplo acima é composto por 45 palavras. A oração iniciada em “para condenar [...]” está subordinada a “pedidos” e, por isso, foi computada na totalidade de CFCV.
DG	b) Além dele, a polícia prendeu <u>Luiz Reinaldo Vieira de Melo, 37 anos, que tem duas passagens na polícia por assalto a banco no Rio Grande do Sul, uma em 1995 e outra em 2000, e o cunhado dele, Régis Lopes Soares, 22 anos.</u>
	O CFCV sublinhado no exemplo acima é composto por 38 palavras, que foi a maior extensão desse tipo de constituinte no DG. No interior desse CFCV sintagma nominal encontramos uma oração relativa que especifica um dos núcleos desse CFCV composto.
JM	c) O Ministério Público da Bahia (MP-BA) encaminhou, ontem, <u>um ofício à Assembleia Legislativa da Bahia (ALBA), solicitando informações sobre os argumentos legais e o valor total pago aos deputados estaduais em decorrência da convocação extraordinária provocada pelo governador Rui Costa (PT), em janeiro, para apreciação de projetos de lei e da Proposta de Emenda à Constituição (PEC) 159/2020, que modifica regras do sistema previdenciário dos servidores públicos do Estado da Bahia aprovada a portas fechada no dia 31 de janeiro, após invasão e tumulto no plenário da Assembleia.</u>
	O trecho acima é composto de 1 frase só com 88 palavras. Dessas 88 palavras, 80 representam o CFCV sublinhado.

Entre os exemplos acima, chama a atenção o último, que é um período composto por 88 palavras organizadas em apenas 1 frase numa sucessão de elementos subordinados. Chama também atenção o fato de ele ter sido extraído das páginas de um jornal considerado popular, como o JM.

Em relação à Amostra 2, o CFCV com maior extensão é o que consta no exemplo (7.38), com 56 palavras. Colocamos abaixo com o número (7.40)

(7.40) O Tribunal a quo negou provimento ao apelo interposto pelo ora agravante, sob o fundamento de que, nas relações de consumo, as cláusulas limitativas de direito serão sempre interpretadas a favor do consumidor, desse modo, ao assim decidir, adotou posicionamento consentâneo com a jurisprudência desta egrégia Corte, que se orienta no sentido de considerar que, em se tratando de contrato de adesão submetido às regras do CDC, a interpretação de suas cláusulas deve ser feita da maneira mais favorável ao consumidor, bem como devem ser consideradas abusivas as cláusulas que visam a restringir procedimentos médicos.

Em relação aos CFCVs oracionais, as Tabelas 7.29 e 7.30 nos apresentam os seguintes dados.

Tabela 7.29 – CFCVs Oracionais por Número de Palavras

TIPO DE CFCV	CORPORA	OCORRÊNCIA POR NÚMERO DE PALAVRAS						
		1-5	6-10	11-20	21-30	31-40	41+	TOTAL
CFCVP	AMOSTRA 2	3	2	4	0	3	2	14
	DG	6	1	5	1	0	0	13
	JM	4	3	2	1	0	0	10
CFCVNP	AMOSTRA 2	9	32	56	25	7	1	130
	DG	24	29	20	3	0	0	76
	JM	21	46	34	15	8	1	125
TOTAL		67	113	121	45	18	4	368

Tabela 7.30 – Percentuais de CFCVs Oracionais por Número de Palavras

TIPO DE CFCV	CORPORA	PERCENTUAIS DE OCORRÊNCIA POR NÚMERO DE PALAVRAS					
		1-5	6-10	11-20	21-30	31-40	41+
CFCVP	AMOSTRA 2	0,26%	0,17%	0,35%	0	0,26%	0,17%
	DG	0,47%	0,08%	0,39%	0,08%	0	0
	JM	0,32%	0,25%	0,16%	0,08%	0	0
CFCVNP	AMOSTRA 2	0,78%	2,78%	4,89%	2,18%	6,1%	0,08%
	DG	1,9%	2,30%	1,59%	0,24%	0	0
	JM	1,71%	3,75%	2,78%	1,22%	0,65%	0,08%

Pela última linha da Tabela 7.29, verifica-se que a maior ocorrência de CFCVs oracionais se dá na faixa de 11 a 20 palavras, seguidos dos que se encontram na faixa de 06 a 10 palavras.

No total de ocorrências de CFCVNPs colocados na última coluna, o DG apresenta quase a metade de ocorrências das sentenças e do JM. Estes dois últimos se assemelham muito na quantidade de CFCVNPs: as sentenças com 130 ocorrências, e o JM com 125.

Conforme a Tabela 7.30, que traz os percentuais de ocorrências em relação à totalidade de CFCVs em cada *corpora*, nas faixas de CFCVs oracionais com mais de 30 palavras, alguns dados se destacam, quais sejam:

- a) A Amostra 2 apresenta um percentual de 6,61% de ocorrências.
- b) O JM apresenta um percentual de 0,73%.
- c) O DG não apresenta nenhuma ocorrência.

Ou seja, as Sentenças apresentam 9 vezes mais complementos verbais com mais de 30 palavras que o JM e o DG.

E a Tabela 7.31 traz dados sobre a distribuição de orações desenvolvidas e reduzidas entre esses constituintes frasais das amostras.

Tabela 7.31 – Tipos de CFCVs Oracionais

TIPOS DE COMPLEMENTOS VERBAIS ORACIONAIS	NÚMERO DE OCORRÊNCIAS			
	Amostra 2	DG	JM	TOTAL
Oração desenvolvida (com a conjunção)	116	52	106	274
Oração reduzida	28	37	29	94
TOTAL	144	89	135	368

Aqui se repete o ocorrido com os constituintes frasais em forma de oração tanto na função de sujeitos como na função de complementos verbais analisados até este momento. Ou seja, a ocorrência desses constituintes na forma de orações desenvolvidas é praticamente três vezes maior que a ocorrência na forma de orações reduzidas. Como dissemos, a presença do conectivo é sempre um facilitador, pois as formas nominais das reduzidas requerem mais trabalho do leitor na busca de um referente para esses verbos.

No Quadro 7.12 apresentamos alguns exemplos de CFCVs oracionais das amostras.

Quadro 7.12 – Exemplos de CFCVs Oracionais das Amostras

AMOSTRA	EXEMPLO
AMOSTRA 2	<p>a) DIANTE DO EXPOSTO, opino pela PARCIAL PROCEDÊNCIA dos pedidos formulados para o fim de: [...]</p> <p>c) condenar a requerida a indenizar danos morais na importância de <u>R\$5.500,00 (cinco mil e quinhentos reais), acrescida de correção monetária pelo IGP-M, a partir da homologação deste parecer (Súmula 362 do STJ), e de juros de 1% ao mês a contar da citação.</u></p> <p>(40 palavras – oração reduzida de infinitivo)</p>
	<p>b) Refiro, ainda, que a finalidade da reparação do dano moral é <u>oferecer compensação ao lesado atenuando seu sofrimento e, quanto ao causador do dano, tem caráter sancionatório para que não pratique mais ato lesivo a personalidade das pessoas.</u></p> <p>(36 palavras – oração desenvolvida com o conectivo que)</p>
	<p>O primeiro exemplo é de um CFCVP composto de 40 palavras, e o segundo, é um CFCVNP composto de 36 palavras. Neste último exemplo, há ainda outras orações hierarquicamente subordinadas a algumas palavras desse constituinte frasal.</p>
DG	<p>c) A hipótese da polícia é de que seja o mesmo bando <u>que cometeu os assaltos a postos de combustíveis no começo da tarde de domingo, só que tripulando um Astra escuro, também roubado.</u></p> <p>(28 palavras - oração desenvolvida com o conectivo que)</p> <p>d) – Soubemos ontem (segunda-feira) <u>que um deles sumiu da vila porque tinha sido agredido pelo patrão do tráfico, justamente por falar sobre o crime e quem tinha ordenado a execução</u> – afirmou o delegado.</p> <p>(26 palavras - oração desenvolvida com o conectivo que)</p>
JM	<p>e) Questionado sobre como essa expansão dos testes está sendo viabilizada nos estados e municípios e o que já havia definido para a Bahia, o Ministério se limitou a informar, por meio de nota, que <u>“ampliou em 869% a capacidade de realização de exames RT-PCR na Rede Nacional de Laboratórios de Saúde Pública”</u></p> <p>(25 palavras - oração reduzida de infinitivo)</p>
	<p>f) A pasta informou <u>que não comenta medidas em análise ou que não sejam públicas</u> e disse <u>que o grupo de monitoramento da crise econômica relacionada à Covid-19, analisa diversas alternativas para reduzir os impactos da pandemia para o setor produtivo e para o setor público, com o objetivo de preservar especialmente a população mais vulnerável.</u></p> <p>(11 e 39 palavras - orações desenvolvidas com o conectivo que)</p>

Em todos os exemplos do Quadro 7.12, os CFCVs oracionais apresentam outras orações hierarquicamente subordinadas a alguns de seus elementos. É o caso, por exemplo, de:

- **oferecer compensação ao lesado**, que se relaciona à forma verbal é no exemplo **b** a Amostra 2.

- **porque tinha sido agredido pelo patrão do tráfico**, que se relaciona à oração **que um deles sumiu da vila**, no exemplo **d** do DG.

- **para reduzir os impactos da pandemia para o setor produtivo e para o setor público**, no exemplo **f** do JM.

Com base nos dados sobre os CFSs e CFCVs de nossas amostras, a partir deste ponto fazemos algumas ponderações sobre o que eles nos permitem concluir.

a) CFSs

Quanto aos CFSs, uma das primeiras observações que fizemos foi a relativa à possibilidade de as sentenças conterem CFSs e CFCVs com maior número de palavras que os jornais por apresentarem menor número de frases e maior número de palavras por frases que os jornais. Nossa observação se confirmou em relação aos CFSs.

Conforme a Tabela 7.21, na faixa de constituintes com 21 ou mais palavras, encontramos 19 ocorrências para as Sentenças, contra 11 no JM e 4 no DG. Essas ocorrências, em termos percentuais relativos à totalidade de CFSs de cada *corpora*, representam os seguintes índices: 1,61% na Amostra 1; 0,76 no JM; e 0,28% no DG. Em termos percentuais, a Amostra 1 apresentou mais que o dobro de ocorrências de CFSs mais extensos que os jornais. Diante disso, podemos afirmar que as Sentenças são mais complexas que os jornais quanto à extensão dos CFSs.

Em relação à ocorrência por tipos de sujeitos, a soma de sujeitos elípticos, pós-verbais e oracionais na Amostra 2 representou 59,81% do total de sujeitos analisados. Nos jornais, a soma das ocorrências desses mesmos tipos de sujeitos gerou os seguintes índices de ocorrência: 40,25% no DG, e 39,55% no JM. Tendo em conta que sujeitos elípticos, pós-verbais e oracionais demandam maior sobrecarga de trabalho, as Sentenças se apresentam mais complexas que os jornais também nesse quesito.

Especificamente em relação aos sujeitos pós-verbais, a maior ocorrência se dá nos textos das Sentenças, com praticamente o dobro de ocorrências do que nos jornais. Em relação à quantidade de palavras, a Amostra 2 repete a ocorrência em mais que o dobro que os jornais: 19,5% contra 7,21% no DG e 9,7% no JM. Esse comportamento reflete o

que ocorreu na totalidade dos sujeitos aqui analisados, conforme expusemos no segundo parágrafo desta seção.

Um fato que nos chamou muito a atenção foi a ocorrência de sujeitos elípticos muito semelhante nos três *corpora*. Ao longo das análises, explicamos as possíveis razões para essas ocorrências nas Sentenças. Quanto aos jornais, temos o estudo de Finatto *et al.* (2011), que analisou textos do DG e do jornal Zero Hora e chegou à conclusão de que a elipse não seria um fator decisivo na avaliação da complexidade textual naquele cenário discursivo. Mas quanto à frequência de sujeitos elípticos, foco do nosso trabalho aqui, podemos supor que as elipses seriam um dos recursos empregados para evitar a repetição de palavras.

Outro dado sobre o qual, ao longo das análises, registramos algumas observações foi o comportamento semelhante da Amostra 2 em relação à Amostra 1 de sentenças. A seguir listamos as situações:

- a. A ordem de ocorrência por tipos de CFSs foi idêntica nas duas amostras.
- b. Os sujeitos oracionais na Amostra 2, tal qual na Amostra 1, aparecem com mais frequência na forma de orações desenvolvidas do que de orações reduzidas e na faixa de 11 a 20 palavras.
- c. Os sujeitos elípticos representados por verbos na 3ª pessoa são os mais frequentes.
- d. Os pronomes relativos foram os mais frequentes entre os CFSs pronominais.

Entendemos que essas situações serviram para comprovar a representatividade das amostras.

A seguir, de modo sintetizado, apresentamos as conclusões a que chegamos com as comparações entre as amostras em termos de complexidade.

Sentenças:

As sentenças se apresentaram mais complexas que os textos dos dois jornais populares em foco nas seguintes situações:

- a. Extensão dos CFSs: em termos percentuais, a Amostra 2 apresentou mais que o dobro de ocorrências de sujeitos mais extensos que os jornais.
- b. Tipos de CFSs: a Amostra 2 apresentou um percentual relativo à soma de sujeitos mais complexos - elípticos, pós-verbais e oracionais – maior que o percentual relativo a essa mesma soma no DG e no JM.
- c. CFSs Pós-Verbais: a Amostra 2 apresentou mais que o dobro de ocorrência de sujeitos pós-verbais que os jornais. Além disso, são mais extensos.
- d. CFSs Elípticos: a Amostra 2 apresentou maior ocorrência de sujeitos elípticos representados por verbos na 3ª pessoa.

- e. CFSs Oracionais: a ocorrência de sujeitos oracionais nas Sentenças superou a dos jornais em mais de 30%.

Jornais DG e JM

E os jornais se apresentaram mais complexos na seguinte situação:

- a. Sujeitos Pré-Verbais: os jornais apresentam sujeitos mais extensos que as sentenças.
- b. Sujeitos Pronominais: os jornais apresentam mais do que o dobro de CFSs pronominais que a Amostra 2.

A maior incidência de CFSs pronominais nos jornais do que nas sentenças pode indicar uma certa desconsideração com as recomendações dos manuais de redação quanto à “pobreza de estilo”. Isso talvez se dê pela preocupação com o público-alvo desses jornais – classes C, D, E, em princípio pessoas menos letradas e com maior dificuldade na leitura. Do nosso ponto de vista, porém, este uso é um dificultador para a leitura. Como já mencionamos, os pronomes demandam maior trabalho no processamento da leitura, porque é necessário encontrar o termo que a ele se refere. Neste caso, então, os jornais se mostram mais complexos que as sentenças.

Além desses dados, cumpre mencionar também que o JM se mostrou mais complexo que o DG em muitos aspectos, principalmente no relativo à extensão das frases e na estrutura do período, com mais elementos subordinados e, conseqüentemente, mais complexos. Conforme referimos, este dado confere com o que consta na Tabela 7.19 em relação ao número de palavras por frase.

b) **CFCVs**

Retomamos aqui nossa observação inicial relativa à possibilidade de as Sentenças conterem CFSs e CFCVs com maior número de palavras que os jornais por apresentarem menor número de frases e maior número de palavras por frases que os jornais. Ela se confirma também em relação aos complementos verbais.

Os dados, em números brutos, das Tabelas 7.27 e 7.29 inicialmente parecem contradizer o que afirmamos no parágrafo anterior. Isso porque o JM, um jornal popular, a partir da faixa de 21 palavras, apresenta 52 ocorrências contra 42 das Sentenças. Entretanto, conforme as Tabelas 7.28 e 7.30, que trazem os percentuais calculados pela relação entre o número de CFCVs encontrados em cada faixa de número de palavras e o total de CFCVs de cada *corpora*, esse quadro se altera.

Em termos percentuais, a Amostra 1 apresenta 11,32% de CFCVs com 21 palavras ou mais, contra 3,91% no JM e 2,25% no DG na mesma faixa. Com base nessa última comparação, podemos, então, afirmar que as Sentenças apresentam CFCVs mais extensos que os dos jornais.

Quanto à forma como se apresentam os CFCVs nas três amostras analisadas, verificamos, pela Tabela 7.26, que a grande maioria está na forma de sintagmas nominais. O mesmo ocorreu com os CFSs. Entretanto, como vimos nos exemplos trazidos, muitos desses constituintes são expandidos por estruturas com vários níveis de subordinação. Tal constituição, como também já mencionamos, demanda maior trabalho na leitura, e, portanto, mais complexo se torna o texto em que se inserem.

O DG, por sua vez, entre as três amostras, foi o que apresentou menor índice de CFCVs sintagmas nominais longos. Nesse quesito, então, em ordem de complexidade, ficaram primeiro os textos das Sentenças, depois os do JM e, por último os do DG.

Entre os CFCVPRs, verificamos na Tabela 7.26, que o DG apresentou maior incidência de pronomes relativos na função de complemento verbal: Amostra 1, 1,4%; JM, 1,4%; e DG, 1,8%. A diferença é pequena, mas é possível relacionar com um dos achados no Estudo-Piloto 1. Nesse estudo, o DG foi o texto que apresentou maior índice de conectivos.

Quanto aos CFCVs oracionais, o DG apresenta quase a metade de ocorrências que se deram nas sentenças e no JM. Além disso, não apresenta nenhuma ocorrência desse tipo de CFCV com mais de 30 palavras, diferentemente do que ocorre com as sentenças e o JM. Este dado mantém o DG na liderança como texto em princípio mais fácil.

Sintetizando as observações sobre os CFCVs encontrados nas amostras analisadas, as Sentenças apresentaram mais constituintes extensos que os jornais. Nesse quesito, ficam como o *corpus* mais complexo entre os três analisados. Em segundo lugar, ficou o JM, e o DG fica como o texto menos complexo.

Quanto ao comportamento semelhante da Amostra 2 em relação à Amostra 1 no caso dos CFCVs, registramos o que segue:

- Ambas as amostras apresentaram mais que o dobro de CFCVNPs do que os CFCVPs: Amostra 1, 6998 CFCVNPs para 3108 CFCVPs; Amostra 2: 758 CFCVNPs para 327 CFCVPs

- O percentual de CFCVPRs também foi aproximado: Amostra 1 - 6,28%; Amostra 2 - 5,44%.

- O percentual de sintagmas nominais e oracionais também foi semelhante: Amostra 1, 81,51% de sintagmas nominais para 11,61% de sintagmas oracionais; Amostra 2, 82,04% de sintagmas nominais para 12,64% de sintagmas oracionais.

- Entre os CFCVPRs representados por pronomes retos, a Amostra 1 teve 2,17% de ocorrência em relação ao total desses constituintes, e a Amostra 2 não teve nenhuma ocorrência.

Tais dados reforçam, do nosso ponto de vista, a representatividade das Amostras.

Concluindo, então, este quarto Estudo-Piloto, numa amostra que julgamos representativa do nosso CE e dos nossos *corpora* de contraste, fica evidente a importância desta pesquisa, principalmente no que se refere a propostas concretas de simplificação sintática. A presença de constituintes frasais longos e com vários elementos subordinados não combina com uma linguagem simples.

Na continuidade desta pesquisa, de modo a ampliar as análises de nossos *corpora* e também para confirmar ou refutar os achados dos estudos até aqui realizados, trazemos dados relativos aos índices de algumas métricas do sistema NILC-Matrix quanto aos nossos *corpora* de contraste.

8 MÉTRICAS DO NILC-METRIX ANALISADAS

Neste capítulo, ampliando as análises até aqui realizadas, apresentamos dados contrastivos entre o CE e os *corpora* de contraste CCA (textos dos jornais Folha de São Paulo) e CCB (jornais populares Diário Gaúcho e Massa), relacionados na seção 5.1 deste trabalho. Os dados relativos a essas análises têm por base os índices de algumas métricas geradas automaticamente pelo sistema NILC-Metrix (Coh-Metrix 3.0, nova versão de 2020).

As métricas que selecionamos estão listadas a seguir. A lista traz o nome da métrica e o grupo ou plano de análise em que ela está inserida no sistema.

- a. Densidade semântica - (Diversidade Lexical)
- b. (#) Média de palavras concretas no texto⁹⁹ - (Psicolinguística)
- c. (#) Média da idade de aquisição das palavras do texto - (Psicolinguística)
- d. Incidência de Adjetivos - (Informações Morfossintáticas de Palavras)
- e. Número de palavras por sentença - (Medidas Descritivas)
- f. Índice Flesch - (Índices de Leitabilidade)
- g. Honoré - (Índices de Leitabilidade)
- h. (#) Distância na árvore de dependências - (Complexidade Sintática)
- i. Número de palavras antes de verbos principais - (Complexidade Sintática)
- j. (#) Proporção de orações relativas em relação à quantidade de orações do texto - (Complexidade Sintática)
- k. Incidência de pronomes - (Informações Morfossintáticas de Palavras)
- l. Incidência de conectivos total - (Conectivos - Complexidade Sintática)
- m. (#) Proporção de orações subordinadas pela quantidade de orações do texto - (Complexidade Sintática)
- n. Proporção de Conjunções Subordinativas em relação a todas as conjunções do texto - (Complexidade Sintática)
- o. (#) Proporção de sujeitos pospostos em relação a todos os sujeitos do texto - (Complexidade Sintática)
- p. (#) Proporção de orações que não estão no formato SVO (sujeito-verbo objeto) em relação a todas as orações do texto - (Complexidade Sintática)

⁹⁹ As métricas assinaladas com (#) são as que constam apenas no sistema Nilc-Metrix, na sua versão mais atual, de 2020 - Coh-Metrix 3.0.

- q. Quantidade média de referentes que se repetem nos pares de sentenças adjacentes do texto - (Coesão Referencial)
- r. Quantidade média de referentes que se repetem nos pares de sentenças do texto - (Coesão Referencial)

À medida que formos apresentando os resultados de cada uma das métricas, explicaremos o que mede cada uma delas e as razões pelas quais foram selecionadas para esta pesquisa. Antes disso, porém, trazemos alguns dados iniciais a respeito dos *corpora* que serão aqui contrastados, tal como fizemos nos Estudos-Pilotos 3 e 4.

Tabela 8.1 – Dados do CE e dos *Corpora* de Contraste

	CE	CCA		CCB	
		ZH	FSP	DG	JM
TOTAL DE TOKENS	542.836	507.075	544.246	496.712	491.667
TOTAL DE TYPES	14.665	28.630	30.333	30.196	27.990
TTR	2,71	5,64	5,57	6,07	5,69
HONORÉ	908,981	1097,684	1049,926	1183,600	1163,219
TOTAL DE FRASES	32.157	29.574	23.559	45.354	28.446
RAZÃO ENTRE FRASE E TOKENS	5,92	5,86	4,32	9,13	5,74
PAL/FRASE	16,706	18,300	20,345	12,096	17,042

Conforme as informações sobre o número de *tokens* na Tabela 8.1, o JM ficou com o menor número (491.667), e o CE com o maior (542.836).

Na linha da TTR, temos os dados da relação entre *types* e *tokens*, que caracteriza a variação lexical de cada *corpus*. Chama a atenção o índice do CE, quase 50% menor que a TTR dos demais *corpora*. Tal dado confere com o que apresentamos nos Estudos-Piloto 1 e 4, em que os textos das sentenças ficaram também com menor variação lexical que os textos dos jornais. Confere também com o fato de o CE ser constituído de textos que veiculam uma linguagem especializada, com muita terminologia e com temáticas também definidas. Tais fenômenos levam à maior repetição de palavras e, conseqüentemente, à menor variação lexical.

Quanto aos textos dos jornais, cumpre referir que uma das orientações dos manuais de redação jornalística é a não repetição de palavras¹⁰⁰. Além disso, a temática que veiculam é bem mais variada que as do CE. São razões possíveis para a maior variação lexical que apresentam.

Logo abaixo da linha relativa à TTR, colocamos os índices da métrica Honoré, que mede a relação entre os *tokens*, os *hapax legomena* (palavras que aparecem uma única vez no texto) e as palavras lexicais, ou palavras de conteúdo, do texto. Considerando as palavras de conteúdo e as palavras que aparecem uma única vez em relação à totalidade de palavras do texto, tem-se uma forma mais equilibrada de avaliar a variação lexical. Quanto maior é o índice desta métrica, mais rico lexicalmente é o texto. Sendo mais rico lexicalmente, ou seja, com maior variação de palavras, mais complexo tende a ser o texto, porque exige que o leitor tenha conhecimento de um repertório maior de palavras.

Pelos dados da Tabela 8.1, em relação a esta métrica, o texto mais rico e, portanto, mais complexo, é do DG. Em segundo lugar, ficaram os textos do JM; em terceiro, o jornal ZH; em quarto, o jornal FSP; e, por último, o CE. Esses dados conferem com os dados da TTR, em que os *corpora* ocupam a mesma posição no *ranking*. Entretanto, é preciso observar que o fato de um texto, como o DG, apresentar maior variação nas palavras pode não significar realmente uma maior complexidade. E o fato de o CE apresentar menor variação lexical pode também não significar, de modo absoluto, uma menor complexidade.

Como vimos no Estudo-Piloto 2, 38,09% das palavras do *corpus* de sentenças estavam ausentes no CorPop¹⁰¹, sendo esse acervo uma referência de “simplicidade” vocabular. E, do *corpus* de textos do DG, foram 23,24% de palavras ausentes no CorPop. **Consumerista, avença, indébito** foram algumas das palavras das sentenças ausentes no CorPop e **açougueiro, alagamento, manicure** foram algumas das palavras do DG ausentes no CorPop. Não é preciso maiores considerações para concluir que, mesmo ausentes no CorPop, palavras como **açougueiro, alagamento, manicure** fazem parte do dia a dia do cidadão brasileiro e provavelmente sejam muito mais conhecidas que **avença, consumerista** e **indébito**. E, nesse ponto, vale sempre relativizar o valor de ponto de referência que se coloca para o CorPop. Afinal, ele, sozinho, na sua amplitude de fontes,

¹⁰⁰ O Diário Gaúcho tem um pequeno manual, de 1995, que foi distribuído apenas para os seus redatores.

¹⁰¹ Lembrando: o CorPop é um *corpus* de referência do Português Popular Escrito, desenvolvido no trabalho de Pasqualini (2018), disponível para consulta em: <http://www.ufrgs.br/textecc/porlexbras/corpop/index.php>.

não resolve a questão multifacetada sobre um dado vocábulo ser potencialmente mais simples ou não. Daí a necessidade de recorrer a outras fontes e recursos.

Na continuidade deste capítulo, vamos trazer índices relativos ao léxico que elucidarão com mais dados o que expusemos no parágrafo anterior.

Além do total de frases apresentados na Tabela 8.1, criamos uma linha para apresentar a relação entre este total de frases e o número de *tokens* em cada *corpora* para termos uma comparação em percentuais. Com isso, acreditamos ter dados mais objetivos sobre os *corpora* quanto a esse aspecto. O fato de o jornal FSP apresentar, em relação aos outros *corpora*, menor razão entre o número de frases e *tokens* confere com o maior número de palavras por frase que também apresenta em relação aos outros *corpora*.

O DG foi o *corpus* com menor número de palavras por frase, o que lhe confere o lugar de menos complexo quanto a este quesito em relação aos outros *corpora*. E o fato de o CE ter ficado o segundo *corpus* com menos palavras por frase pode ser atribuído à grande quantidade de frases nominais existentes nos cabeçalhos das ementas jurisprudenciais reproduzidas nas sentenças. Essas ementas jurisprudenciais são referências a outros julgamentos sobre o mesmo tema e são utilizadas para fundamentar a decisão adotada em cada caso. No exemplo (8.1), reproduzimos o cabeçalho de uma ementa jurisprudencial extraída do CE.

(8.1) APELAÇÃO CÍVEL. RESPONSABILIDADE CIVIL. INSCRIÇÃO EM ÓRGÃO DE RESTRIÇÃO DE CRÉDITO. DÍVIDA EXISTENTE. DANO MORAL NÃO VERIFICADO. EXERCÍCIO REGULAR DO DIREITO.

Nas frases nominais em caixa alta do cabeçalho acima, a mais longa tem 7 palavras. cremos, então, que essa pode ser essa uma razão importante para o índice em relação ao número de palavras por frases nas sentenças ser o segundo mais baixo entre os *corpora*.

Feita esta exposição inicial sobre os dados dos *corpora* aqui analisados, passamos a apresentar os índices relativos a cada uma das métricas escolhidas. Iniciaremos pelas métricas que medem aspectos relativos ao léxico. Ainda que algumas sejam consideradas, pelo sistema NILC-Matrix, descritivas ou morfossintáticas, como é o caso da Incidência de Adjetivos, e outras psicolinguísticas, como a Idade de Aquisição da Palavras, nós as agrupamos como métricas de ordem lexical.

8.1 DENSIDADE SEMÂNTICA

Adotamos esta métrica, pois ela considera o número de palavras de conteúdo (lexicais – substantivo, adjetivo, advérbios e verbos) dividido pelo número de palavras funcionais (gramaticais – preposições, conjunções, numerais e artigos). Segundo Silva (2018, p. 121),

[o] índice de densidade semântica indica o quão complexo um texto é em termos de conteúdo e, a partir desse número, o quanto de vocabulário é necessário para a interpretação dos componentes do texto. Ou seja, textos que tenham índices altos de densidade semântica são aqueles que exigem maior conhecimento previamente adquirido dos leitores e que exijam maior esforço cognitivo para interpretar a significação não só das palavras, mas também da relação entre as palavras.

Conforme vimos no Estudo-Piloto 2, os textos do CE apresentam muitas terminologias e muitas palavras “potencialmente complexas”. Por isso, julgamos pertinente incluir essa métrica em nossas análises, para verificar se os dados dela advindos confirmam os dados relativos às comparações lexicais que realizaremos entre os *corpora* de nossa pesquisa.

A Tabela 8.2 apresenta os dados desta métrica em nossos *corpora*. Quanto mais alto o índice, mais complexo é o texto, porque apresenta mais palavras de conteúdo.

Tabela 8.2 – Densidade Semântica

CORPUS ANALISADO	FSP	ZH	CE	JM	DG
ÍNDICE	1,4976	1,6003	1,6608	1,7567	1,9515
ORDEM DE DIFICULDADE	1 (mais fácil)	2	3	4	5 (mais difícil)

Os números colocados na última linha da Tabela 8.2 representam a ordem de dificuldade de cada *corpora* em relação aos demais. O número 1 representa o mais fácil, e, à medida que o número aumenta, aumenta também a dificuldade. Assim, o número 5 fica como o mais difícil.

No Estudo-Piloto 2, as sentenças, com o índice de 1,648, se mostraram mais complexas que o DG, com o índice de 1,533 quanto à densidade semântica. Aqui, conforme a Tabela 8.2, as sentenças ficaram mais fáceis que os dois jornais populares analisados: o DG, primeiro mais difícil, e o JM, segundo mais difícil quanto a esta métrica.

Este resultado do DG e do JM parece ser compatível com o resultado que apresentaram em relação à TTR (relação entre *types* e *tokens*). O DG foi o *corpus* que apresentou maior TTR, ou seja, que apresentou maior variação lexical, e o JM o segundo.

Entretanto, o fato de o CE ficar em terceiro lugar quanto à densidade semântica não confere com o resultado da sua TTR (último colocado).

É preciso, porém, lembrar que a TTR estabelece uma relação entre todos os tipos de palavras, e a densidade semântica procura distinguir a proporção em que as palavras relacionadas – de conteúdo e funcionais – aparecem no texto. Nessas relações, podem ocorrer variações. É preciso também ponderar o quanto essas palavras de conteúdo que aparecem em maior quantidade nos jornais são ou não “potencialmente complexas”, como vimos no Estudo-Piloto 2 e exemplificamos com os casos de **avença**, **consumerista** e **indébito**, no CE, e **açougueiro**, **alagamento**, **manicure**, no DG.

Por essa razão, temos um capítulo para comparações lexicais dos *corpora* aqui analisados e também com mais outros *corpora*, conforme especificado no capítulo relativo aos procedimentos metodológicos.

Neste momento, porém, dando continuidade às análises advindas do sistema NILC-Metrix, apresentamos métricas relativas a algumas características psicolinguísticas das palavras. Acreditamos que elas poderão trazer, juntamente com a densidade semântica e a TTR, uma melhor ponderação das particularidades lexicais de nossos *corpora*.

8.2 MÉDIA DE PALAVRAS CONCRETAS NO TEXTO

Esta métrica mede o valor de concretude das palavras de conteúdo do texto. Quanto mais palavras concretas um texto apresenta, mais fácil ele tende a ser. Assim ocorre, porque palavras concretas referem-se a coisas, objetos dos quais se pode formar uma imagem mental, que se possa ver ou tocar concretamente.

Vejam os seguintes exemplos:

(8.2) Meu **cachorro** ficou doente.

(8.3) A **doença** do meu cachorro me causou **ansiedade**.

A palavra **cachorro**, no exemplo (8.2), é uma palavra concreta. É possível se fazer uma imagem mental a respeito dela. Já **doença** e **ansiedade**, no exemplo (8.3), não são palavras concretas. É difícil ter uma imagem mental dessas palavras, bem como é impossível tocar o referente delas. É o que ocorre com **avença**, **consumerista** e **indébito**, no CE. Já **açougueiro**, **alagamento** e **manicure**, no DG, são palavras que permitem uma imagem mental a seu respeito.

A medida de concretude das palavras realizada pelo sistema NILC-Metrix se dá com base no DELAF¹⁰², um dicionário computacional do português brasileiro (PB) do sistema UNITEX. O UNITEX, por sua vez, “é um conjunto de *softwares* que permite processar os textos em línguas naturais utilizando recursos linguísticos. Esses recursos se apresentam na forma de dicionários eletrônicos, de gramáticas e tabelas de léxico-gramática” (PAUMIER, 2007, p. 4).

Janczura, Castilho, Rocha, Van Erven e Huang (2007), em artigo sobre normas de concretude das palavras, afirmam, com base em pesquisas de outros autores, que, na leitura, palavras concretas são mais fáceis de serem compreendidas e memorizadas do que as abstratas. Entre vários estudos citados, trazem dados como os que seguem:

- a. Os indivíduos respondem mais rapidamente a questionamentos que envolvam sentenças concretas do que abstratas.
- b. Na aprendizagem da língua nativa, palavras concretas são adquiridas mais cedo e com mais facilidade que as abstratas.
- c. Na aprendizagem de uma segunda língua, palavras concretas são mais fáceis de adquirir e de serem memorizadas.

Diante disso e considerando que a terminologia da área do Direito é permeada por abstrações, julgamos importante utilizar esta métrica na comparação de nossos *corpora*. Uma de nossas hipóteses de pesquisa é a de que a alta presença de terminologia é um elemento propulsor da complexidade dos textos do nosso CE. Por isso, cremos que esta métrica pode nos ajudar a verificar em que medida isso realmente ocorre nos textos aqui analisados.

A Tabela 8.3 apresenta os dados relativos à concretude das palavras de nossos *corpora*. Quando maior o valor desta métrica, mais fácil tende a ser o texto, porque apresenta maior número de palavras concretas.

Tabela 8.3 – Média de Palavras Concretas no Texto

CORPUS ANALISADO	DG	JM	ZH	CE	FSP
ÍNDICE	4,3487	4,2803	4,1401	3,4620	3,4059
ORDEM DE DIFICULDADE	1 (mais fácil)	2	3	4	5 (mais difícil)

Como vemos na Tabela 8.3, o texto que apresenta mais palavras concretas e, portanto, é o mais fácil é o do DG, seguido do JM. E o mais difícil é o do jornal FSP, seguido dos textos do CE. O jornal ZH ficou no meio-termo.

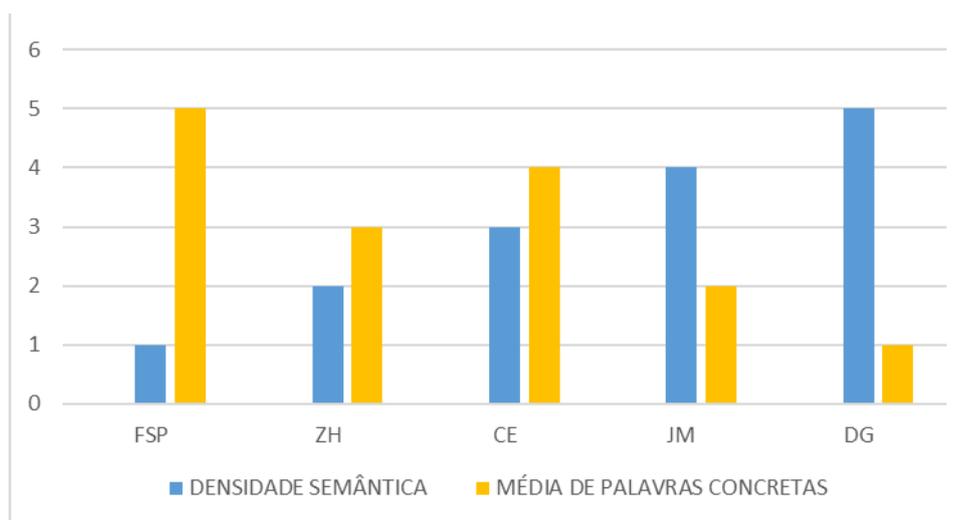
¹⁰² Finatto, Vale e Laporte (2019), em artigo em que relatam um experimento de verificação do vocabulário de jornais populares pelo DELAF-PB, trazem dados mais detalhados sobre esse dicionário.

O resultado relativo ao DG e ao JM é o esperado, visto que são jornais populares e, como já mencionamos várias vezes, têm como público-alvo um leitor menos escolarizado. O fato de o jornal FSP ter ficado como mais complexo que o CE surpreende um pouco, mas também se justifica, considerando também o tipo de jornal que é e o público a que se destina.

Mesmo tendo ficado em segundo lugar no *ranking* dos mais difíceis em relação às palavras concretas, este resultado parece confirmar nossa hipótese de que a alta presença de terminologia e de palavras abstratas é um fator propulsor para complexidade dos textos das sentenças.

A seguir, estabelecendo uma relação desta métrica com a Densidade Semântica, medida em que o DG ficou como o mais difícil, apresentamos o Gráfico 8.1 a seguir.

Gráfico 8.1– Densidade Semântica x Palavras Concretas



O eixo vertical do Gráfico 8.1 representa os níveis de dificuldade dos textos: o número 1 representa o mais fácil, e o 5 o mais difícil. Conforme se vê, então, no Gráfico 8.1, as duas medidas – densidade semântica e *concretude das palavras* – são inversamente proporcionais em relação aos primeiros e últimos colocados. O Jornal FSP, mais fácil em relação à densidade semântica, ficou como o mais difícil em relação à média de palavras concretas em seu texto. Ou seja, é um texto que apresenta poucas palavras de conteúdo, mas muitas abstratas. E com o DG, ocorre exatamente o inverso: apresenta muitas palavras de conteúdo, mas poucas abstratas.

Esta comparação justifica a observação que fizemos em relação à complexidade das palavras de conteúdo que aparecem no texto nos comentários sobre a métrica

Densidade Semântica. Também reforça que o léxico do DG, apesar de ser mais variado, é mais concreto e, portanto, mais fácil.

8.3 (#) MÉDIA DA IDADE DE AQUISIÇÃO DAS PALAVRAS DO TEXTO

Esta métrica procura verificar a idade em que a palavra foi adquirida pelo falante. Segundo o Manual do NILC-Metrix¹⁰³,

idade de aquisição é uma característica psicolinguística das palavras de conteúdo e representa o intervalo de idade em que a palavra foi adquirida. Os valores variam de 1 a 7 e quanto menor o valor, menor a idade de aquisição. Os valores correspondem a faixas de idade. 1 (de 0 a 2 anos); 2 (de 3 a 4 anos), 3 (de 5 a 6 anos), 4 (de 7 a 8 anos), 5 (de 9 a 10 anos), 6 (de 11 a 12 anos), 7 (13 anos ou mais).

São computadas as palavras de conteúdo do texto e relacionadas com um repositório psicolinguístico¹⁰⁴ de 26.874 palavras. Segundo o Manual do NILC-Metrix, esse repositório psicolinguístico, com valores de concretude, familiaridade, idade de aquisição e imageabilidade, está sujeito a vieses porque foi construído automaticamente. Esses dados foram anotados por psicolinguistas e psicólogos.

A Tabela 8.4 apresenta os dados relativos a esta métrica. Quanto maior o valor da média, mais complexo é o texto, porque maior é a idade em que a palavra é adquirida pelo falante.

Tabela 8.4 – Média da Idade de Aquisição das Palavras do Texto

CORPUS ANALISADO	DG	JM	ZH	FSP	CE
ÍNDICE	4,4886	4,7188	4,9393	4,9733	5,6428
ORDEM DE DIFICULDADE	1 (mais fácil)	2	3	4	5 (mais difícil)

Pelo que se vê na Tabela 8.4, os textos dos jornais populares são os que apresentam palavras cuja idade de aquisição é menor, e o CE e o jornal FSP se mostram mais difíceis nesse quesito, com médias que levam a uma idade maior de aquisição das palavras. O jornal ZH manteve-se no meio termo, ocupando o terceiro lugar, tal qual na métrica das palavras concretas.

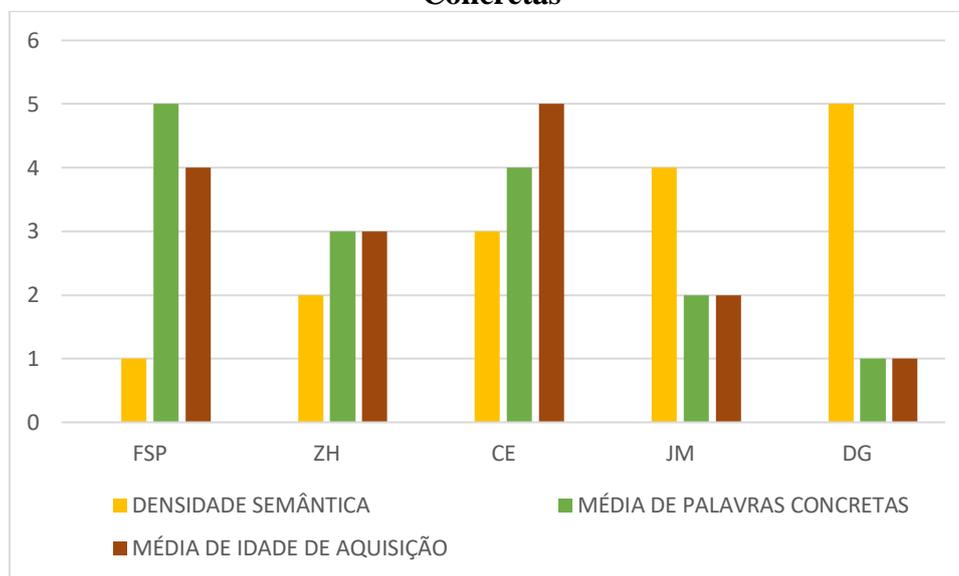
¹⁰³ Disponível em: http://fw.nilc.icmc.usp.br:23380/metrixdoc#medidas_descritivas

¹⁰⁴ O repositório está disponível em:

http://143.107.183.175:21380/portlex/index.php/en/?option=com_content&view=article&layout=edit&id=23

Agregando estas médias à comparação entre Densidade Semântica e Concretude das Palavras, temos Gráfico 8.2 a seguir.

Gráfico 8.2 – Densidade Semântica x Idade de Aquisição x Palavras Concretas



Aqui também o eixo vertical representa a dificuldade dos textos, conforme explicamos no Gráfico 8.1. Pelo Gráfico 8.2 acima, é possível verificar melhor a uniformidade no comportamento dos jornais DG e JM quanto às três métricas aqui analisadas. Ambos são mais fáceis quanto à quantidade de palavras concretas e quanto à idade de aquisição das palavras. Mesmo que sejam mais densos semanticamente, ou seja, que apresentem mais palavras de conteúdo do que funcionais, essas palavras não são tão complexas como as que as métricas revelam em relação aos outros *corpora*.

O CE ficou como o terceiro mais denso semanticamente, com palavras cuja idade de aquisição é a maior em relação a todos os *corpora* e o segundo em relação a palavras abstratas. O jornal FSP foi o menos denso semanticamente, entretanto as palavras que apresenta são as mais abstratas entre todos os *corpora* analisados e com idade de aquisição também superior aos dos demais, perdendo só para o CE. Ou seja, o CE e a FSP são os que, por essas métricas, apresentam palavras “potencialmente mais complexas”.

8.4 INCIDÊNCIA DE ADJETIVOS

Esta métrica mede a proporção de adjetivos em relação a todas as palavras do texto. Em sua pesquisa de Mestrado, Silva (2018) concluiu que a redução do número de

adjetivos no texto diminui a sua complexidade. Diante disso, selecionamos tal métrica para verificar como se comportam nossos *corpora* em relação aos adjetivos.

A Tabela 8.5 apresenta os dados relativos a esta métrica. Quanto maior o valor, maior a incidência de adjetivos e, portanto, mais complexo é o texto.

Tabela 8.5 – Proporção de Adjetivos em relação às demais Palavras do Texto

CORPUS ANALISADO	DG	JM	ZH	FSP	CE
ÍNDICE	0,0352	0,0455	0,0470	0,0485	0,0518
ORDEM DE DIFICULDADE	1 (mais fácil)	2	3	4	5 (mais difícil)

Pelos dados da Tabela 8.5, o texto mais complexo em relação à incidência de adjetivos é o CE, o que é um resultado esperado. Na seção 4.2, tratamos sobre a probabilidade de determinados traços linguísticos variarem em relação a textos típicos de contextos comunicativos específicos. Segundo Hoffmann (2015c), os substantivos e adjetivos representam em torno de 60% do vocabulário dos textos especializados, porque é preciso denominar os conceitos, os objetos e processos especializados. É o que ocorre na linguagem jurídica. Por exemplo, para especificar a área do Direito de que se trata é preciso qualificar com um adjetivo: Direito **Civil**, Direito **Penal**, Direito **Tributário**. E assim se dá em várias outras situações nesse e em também em outros âmbitos especializados.

Além disso, cabe lembrar que adjetivos são palavras abstratas e, portanto, sua presença diminui a concretude do texto, tornando-o, por isso, mais complexo. Como se não bastasse essa característica dos adjetivos, a dificuldade aumenta ainda mais quando se resolve empregar adjetivos pouco comuns, como o que ocorre no seguinte exemplo extraído do CE:

- (8.4) Considerando a exclusão dos recibos de fls. 40 e 43 (notas fiscais das lentes intraoculares), e somando os recibos de fls. 41/42 e 45/46, tem-se o total de R\$ 9.000,00, justamente o valor apontado pela demandante na peça **isagógica**.

O adjetivo **isagógico** não consta nos *corpora* de contraste aqui analisados nem no CorPop. E a expressão **peça isagógica** faz referência ao pedido inicial de uma ação, que tem o nome técnico de **petição inicial**. O redator quis impressionar com um adjetivo diferente, mas acabou sendo impreciso tecnicamente. Isso porque a petição inicial não é

uma simples introdução nem o documento principal do processo, que seria o significado de **isagógico**.

Os jornais populares, DG e JM, seguiram também a lógica esperada. Uma vez que a presença de adjetivos torna o texto mais complexo, e considerando o público-alvo desses jornais, é coerente que empreguem menos qualificativos em seus textos. Os jornais de referência, conseqüentemente, também seguiram a lógica esperada, ficando o jornal FSP como o segundo mais difícil no *ranking*.

Um dado que chama a atenção nessa tabela é a diferença entre os índices de cada *corpus*. Entre o JM, o jornal ZH e o jornal FSP, há uma diferença de 15 centésimos. Mas entre o DG e JM, que lhe segue, há diferença de quase 200 centésimos. Isso acentua mais a diferença do DG nesse quesito em relação aos demais *corpora*.

8.5 ÍNDICE FLESCH

Como já referimos, este índice avalia a complexidade da leitura com base na relação entre tamanho das frases e tamanho das palavras. Ele tem sido bastante utilizado em diferentes estudos, com bons resultados, apesar da superficialidade dos elementos que considera. A Tabela 8.6 apresenta os dados de nossos *corpora* relativos ao Índice Flesch (IF).

Tabela 8.6 – Índice Flesch do CE e *Corpora* de Contraste

CORPUS ANALISADO	DG	JM	CE	ZH	FSP
ÍNDICE	56,7062	49,2196	42,9666	41,3630	40,9373
ORDEM DE DIFICULDADE	1(mais fácil)	2	3	4	5(mais difícil)

Quanto mais alto o valor do IF, mais fácil é o texto. O DG ficou como o texto mais fácil, em coerência com os dados que apresentou em relação ao número de palavras por frase. Em segundo lugar, ficou o JM, que, em relação ao número de palavras por frase, ficou em terceiro.

Em terceiro, ficou o CE, que, em relação ao número de palavras por frase, ficou em segundo.

Os jornais ZH e FSP mantiveram suas posições no *ranking* do número de palavras por frase e quanto ao Índice Flesch. O jornal ZH ficou em 4º lugar, e o jornal FSP em 5º lugar, como o mais difícil entre todos os *corpora*.

Como já comentamos, o IF não pode ser tomado isoladamente, pois nem sempre uma palavra com muitas letras e/ou sílabas é mais difícil que uma palavra com menos letras e/ou sílabas. E nem sempre uma frase mais longa será mais difícil que uma frase mais curta. É o que podemos ver no Quadro 8.1, com exemplos extraídos de nossos *corpora* e relativos ao tamanho das palavras.

Quadro 8.1 – Exemplos do CE e do DG

(8.5) Os autos emprestam verossimilhança à tese da autora de que não contratou da ré os serviços que deram azo à negatificação do seu nome junto aos órgãos de proteção ao crédito. (CE)
A palavra azo , que significa <i>motivo, razão</i> , tem apenas 3 letras. Ela ocorre 7 vezes no CE, mas não consta em nenhum dos <i>corpora</i> de contraste aqui analisados, como também não ocorre no CorPop.
(8.6) Responsabilidade será dos volantes. (DG)
A palavra responsabilidade , em destaque no exemplo acima, tem 16 letras. Ela ocorre 55 vezes no DG, 91 no jornal FSP, 41 no JM, 80 no jornal ZH, 669 no CE e 57 no CorPop.

De todo modo, conforme comentamos no Estudo-Piloto 1, o IF é considerado válido porque ajuda a indicar a potencial complexidade do texto. Além disso, segundo Scarton e Aluísio (2010), foi a primeira métrica de inteligibilidade adaptada para o português que incorporou o conceito de séries escolares brasileiras.

A partir daqui passamos a trabalhar com métricas que colocamos no Grupo da Sintaxe. Apesar de muitas serem agrupadas de outra forma pelo Sistema NILC-Metrix, tal qual fizemos em relação ao Grupo do Léxico por nós criado, nós a colocamos no mesmo conjunto porque as influências que exercem no texto se inter-relacionam.

8.6 DISTÂNCIA NA ÁRVORE DE DEPENDÊNCIAS

Esta métrica calcula a distância entre as palavras que se relacionam na superfície textual. Conforme já mencionamos, grandes distâncias entre palavras que se relacionam

sobrecarregam a memória de trabalho e, conseqüentemente, dificultam a compreensão leitora.

Como analisamos a extensão dos constituintes frasais relacionados ao verbo nos Estudos-Pilotos 3 e 4, julgamos importante trazer os dados relativos a esta métrica para compararmos o CE em relação aos *corpora* de contraste, os jornais ZH e FSP e também o DG e o JM. A ideia é verificar se os dados advindos da análise computacional por meio dessa ferramenta confirmam ou não as análises manuais que fizemos em relação aos CFSs e CFCVs nos Estudos-Pilotos 3 e 4.

A Tabela 8.7 apresenta os dados dessa métrica relativos ao nosso CE e aos nossos *corpora* de contraste. Quanto maior o índice, maior é a distância na árvore de dependência, e, conseqüentemente, mais complexo tende a ser o texto.

Tabela 8.7 – Distância na Árvore de Dependência

CORPUS ANALISADO	DG	ZH	JM	CE	FSP
ÍNDICE	40,8411	40,84113	41,4493	41,60963915	47,95926
ORDEM DE DIFICULDADE	1 (mais fácil)	2	3	4	5 (mais difícil)

Em relação à Distância na Árvore de Dependência, os textos mais fáceis são os do DG. Em segundo lugar, ficaram os textos do ZH; em terceiro, os do JM; em quarto, os do CE; e por último foram os da FSP.

Essa posição do DG é compatível com os dados encontrados no Estudo-Piloto 4. Entre as amostras de textos do DG, do JM e das sentenças, o DG ficou com:

- a) mais frases;
- b) menos palavras por frase;
- c) menos CFSs e CFCVs oracionais;
- d) mais sujeitos pré-verbais;
- e) mais CFSs e CFCVs pronominais;
- f) mais CFSs e CFCVs na faixa de 1 a 5 palavras.

Se o DG ficou com mais frases e menos palavras por frases, é esperado que a distância de dependência entre as palavras seja menor. Além disso, o DG apresentou mais sujeitos pronominais, sujeitos na faixa de 1 a 5 palavras e menos sujeitos oracionais. Essas ocorrências também levam a uma menor distância de dependência entre as palavras.

O jornal FSP ficou mais complexo que os textos do CE quanto a esse aspecto. O resultado do jornal FSP é compatível com o índice relativo ao número de palavras por frase, em que também ficou como o mais difícil.

O CE ficou como o segundo mais difícil. E aqui cabe lembrar que o CE ficou como o segundo colocado em relação ao número de palavras por frase. Mesmo assim, a distância de dependência entre as palavras que se relacionam ficou maior que a dos outros *corpora*, com exceção do jornal FSP.

Nesse sentido, no Estudo-Piloto 4, em comparação com o DG e o JM, as sentenças apresentaram maior ocorrência de constituintes frasais nas faixas superiores a 30 palavras. Entre os complementos verbais, a diferença ficou em quase 50% a mais que o JM, e o DG não teve nenhuma ocorrência.

A posição do JM como mais difícil que o jornal ZH chama a atenção. O JM é considerado um jornal popular e deveria, por isso, ter uma estrutura frasal menos complexa. Apesar disso, os dados relativos a esta métrica conferem com os achados nos Estudos-Pilotos 1 e 4. No Estudo-Piloto 4, o JM apresentou sintagmas nominais com extensão superior aos das sentenças, e, no Estudo-Piloto 1, foi o segundo colocado em relação a parágrafos com maior extensão. Tais comportamentos podem ser razões para a maior distância na árvore de dependência entre as palavras apresentada pelo JM.

8.7 NÚMERO DE PALAVRAS ANTES DE VERBOS PRINCIPAIS.

Esta métrica mede a quantidade média de palavras antes dos verbos principais das orações principais das sentenças. E a quantidade de palavras antes dos verbos principais está, conforme visto no Estudo-Piloto 1, relacionada à memória de trabalho. Quanto mais palavras antes do verbo, mais memória de trabalho é exigida do leitor, o que dificulta a compreensão.

Nesse Estudo-Piloto 1, os jornais populares ficaram como mais complexos que as Sentenças do CE. Tal resultado nos gerou uma certa perplexidade e curiosidade em verificar se isso se confirmaria com *corpora* maiores. Além disso, no Estudo-Piloto 3, quando analisamos os sujeitos pré-verbais, verificamos que muitos se apresentavam delimitados por orações encaixadas, o que aumenta o número de palavras antes do verbo principal e, conforme observamos, é um dificultador da leitura.

A Tabela 8.8 apresenta os dados de nossos *corpora* em relação a esta métrica. Quanto maior o índice, mais palavras antes dos verbos principais e, portanto, mais complexo é o texto.

Tabela 8.8 – Número de Palavras antes dos Verbos Principais

<i>CORPUS</i> ANALISADO	CE	FSP	DG	JM	ZH
ÍNDICE	1,5229	1,6501	4,4449	4,7149	5,9124
ORDEM DE DIFICULDADE	1(mais fácil)	2	3	4	5(mais difícil)

Conforme os dados da Tabela 8.8, o *corpus* com menor número médio de palavras antes dos verbos principais e, portanto, o mais fácil, é o CE. E o *corpus* mais difícil, ou seja, o que apresenta mais palavras antes dos verbos principais é o do ZH. No Estudo-Piloto 1, os textos mais difíceis quanto a esse aspecto foram os do JM, segundo mais difícil na análise atual, e os das sentenças sem a as ementas jurisprudenciais.

Nesse Estudo-Piloto 1, criamos dois *corpora* de sentenças: um com as ementas jurisprudenciais e outro sem. A ideia foi verificar em que medida a presença de muitas frases nominais, como ocorre nos cabeçalhos das ementas jurisprudenciais, afetaria nas medidas trabalhadas. E nesse caso alterou. As sentenças com a presença de ementas jurisprudenciais ficaram mais acessíveis, com índice menor, que as sentenças sem as ementas. E a explicação que encontramos para tal fenômeno foi o fato de as ementas conterem muitas frases nominais e palavras isoladas por ponto, o que é considerado uma sentença. Desse modo, havendo mais sentenças sem verbos, esse índice tende a diminuir.

Na análise atual, não criamos um *corpus* de sentenças sem ementas. Então, cremos que uma das explicações para este fenômeno está no fato de haver mais sentenças sem verbos, o que leva a diminuir o número de verbos e, conseqüentemente, a contagem de palavras antes desses verbos.

De todo modo, a posição do DG e a do jornal FSP, apesar de manter o ocorrido no Estudo-Piloto 1 entre os textos desses dois jornais (o DG ficou com mais palavras antes dos verbos principais que o jornal FSP), surpreende. Em relação à métrica **Distância na Árvore de Dependência**, o DG ficou como o texto mais acessível. Ao analisarmos esse comportamento do DG, julgamos que ele era pertinente com os dados do Estudo-

Piloto 4, segundo o qual o DG apresentou mais frases, menos palavras por frase, mais sujeitos pré-verbais, mais constituintes frasais pronominais e na faixa de 1 a 5 palavras.

Assim, considerando essas características do DG, é de se esperar que haja menos palavras antes dos verbos principais. E o jornal FSP, que apresenta mais palavras por frase que todos os demais *corpora*, ficou ainda mais fácil que o DG em relação ao número de palavras antes dos verbos principais das orações. Uma das possíveis razões para este fenômeno pode estar no fato de muitas formas nominais dos verbos serem empregadas como nomes e serem interpretadas pela ferramenta como verbos. É o caso das sentenças judiciais, que apresentam muitos participios que denominam atores e entidades do processo. Temos, por exemplo, situações como **demandado/demandada; apelado/apelada, requerido/requerida, inominado (recurso inominado), pedido, enunciado**, etc. O participio “demandada”, inclusive, foi alvo de análise equivocada pelo Parser Palavras, conforme referimos no Estudo-Piloto 3.

No Quadro 8.2, a seguir, apresentamos alguns exemplos extraídos do CE que ilustram essas situações.

Quadro 8.2 – Exemplos do CE

<p>(8.7) PEDIDO</p> <p>Do total de 1477 ocorrências, apenas 1 foi como verbo. Nas demais, como substantivo.</p> <p>Como verbo:</p> <ul style="list-style-type: none"> • Em contestação, a Ré alega: a) preliminarmente: nada. b) meritoriamente: que o procedimento pedido pela Autora não foi coberto em razão da limitação geográfica, pois o plano possui cobertura regional [...]. <p>Como substantivo (demais ocorrências):</p> <ul style="list-style-type: none"> • Ainda, afirma que o Sr. XXXXXXXX já havia feito pedido de ligação da luz e que a CEEE já deveria ter atendido a tal solicitação.
<p>(8.8) SENTIDO</p> <p>Entre as 579 ocorrências de sentido, não há sequer 1 que seja como verbo.</p> <p>Como locução adverbial:</p> <ul style="list-style-type: none"> • A temática presente nos autos já foi vastamente aventada pela jurisprudência, que firmou entendimento no sentido da legalidade do faturamento na forma efetivada pela demandada. • As alegações da ré, no sentido de que o autor estaria em alta velocidade ou falando ao telefone não encontram amparo em qualquer elemento [...].
<p>(8.9) REQUERIDO</p>

Entre as 199 ocorrências de **requerido**, apenas 29 são como verbo.

Como verbo:

- Nada sendo **requerido**, baixe-se.

Como substantivo:

- Não há, pois, como condenar **o requerido** a indenizar a autora relativamente à contratação de seu procurador.

Diante disso, cremos que as medidas extraídas por essa métrica não expressam com exatidão o comportamento de nossos *corpora* relativamente ao número de palavras antes dos verbos principais, especialmente quanto ao CE.

8.8 INCIDÊNCIA DE PRONOMES

Esta métrica mede a proporção de pronomes em relação à quantidade de palavras do texto. Nos Estudos-Pilotos 3 e 4, os constituintes frasais representados por pronomes não foram os de maior frequência. Mesmo assim, julgamos importante trazer os dados relativos a esta métrica porque o emprego dos pronomes é tido, nas pesquisas sobre leitura e letramento, como um dificultador da leitura. Isso acontece, como já explicamos, porque a busca pelo referente do pronome pode comprometer a fluência da leitura, principalmente quando o pronome puder se referir a mais de um elemento do texto.

A Tabela 8.9 apresenta os dados relativos à incidência de pronomes. Quanto mais alto o valor desta métrica, maior a incidência de pronomes no texto e, conseqüentemente, mais difícil é esse texto.

Tabela 8.9 – Incidência de Pronomes

CORPUS ANALISADO	CE	JM	ZH	DG	FSP
ÍNDICE	0,0406	0,0413	0,0457	0,0468	0,05916
ORDEM DE DIFICULDADE	1(mais fácil)	2	3	4	5(mais difícil)

Conforme se vê na Tabela 8.9, os textos mais fáceis foram os do CE, JM e ZH, nessa ordem.

No Estudo-Piloto 4, o DG foi o *corpus* que apresentou mais constituintes frasais pronominais, resultado que confere com a atual métrica. O DG, segundo mais difícil neste quesito, perde aqui apenas para o jornal FSP, que não foi analisado no Estudo-Piloto 4.

Quanto ao jornal FSP, cabe lembrar que o perfil de leitor a que se destina são pessoas em princípio com maior escolaridade que os leitores do DG. A frequência maior de pronomes para esse perfil de leitor pode não ser problema, e por isso pode não haver maiores preocupações por parte da redação desse jornal quanto ao uso desse recurso linguístico em seus textos.

É preciso salientar, porém, que os índices aqui apresentados foram em relação ao conjunto total de textos de cada *corpus* e em relação à totalidade de pronomes neles existentes. Nesse conjunto pode haver pronomes que não exerçam função de constituinte frasal. É o caso dos “pronomes adjetivos”, que atuam como modificador de um substantivo. A seguir, no Quadro 8.3, apresentamos alguns exemplos extraídos do DG:

Quadro 8.3 – Exemplos do DG

<p>(8.10) Até ontem à tarde, o segurança – cujo nome foi mantido em sigilo a pedido da família – estava hospitalizado.</p> <ul style="list-style-type: none"> • Aqui o pronome cujo está no CFS cujo nome, mas com a função de adjunto adnominal.
<p>(8.11) Sua carreira terminou há algum tempo, desde que as lesões o impediram de permanecer entre os melhores tenistas do mundo.</p> <ul style="list-style-type: none"> • Neste exemplo, o pronome algum está modificando o substantivo tempo, numa expressão adverbial.

Mesmo que, por esta métrica, não só os pronomes que atuam como constituinte frasal sejam computados, o CE e o JM mantiveram o *ranking* obtido nos Estudos-Pilotos 3 e 4. O CE foi o que apresentou menos ocorrências de constituintes frasais pronominais, principalmente os pronomes pessoais retos. No Estudo-Piloto 3, houve 16 ocorrências, e no Estudo-Piloto 4, nenhuma. Considerando esses fenômenos, naquele momento observamos que eles poderiam ser indícios de menor complexidade textual das sentenças no que se refere ao emprego de pronomes.

Nesses estudos-pilotos referimos que uma das possíveis causas para a pouca frequência de pronomes no CE estaria no fato de os atores e as entidades do processo judicial serem normalmente colocados no texto com a denominação técnica que recebem. No início do relatório, normalmente se coloca o nome da pessoa que ingressa com a ação judicial, o que ela pede, e contra quem ela pede algo. Depois disso, em todas as demais ocasiões em que se faz referência a essa pessoa, normalmente se usa a designação técnica

que se dá a ela, por exemplo, autor/autora, requerente. E isso se comprova pelos seguintes dados:

- **Autora** é a palavra lexical com maior ocorrência no CE: 2764 vezes. No jornal FSP ocorre 5 vezes; no jornal ZH, 2; no JM, 4; e no DG, 3.

- **Autor** é a palavra lexical em quarto lugar nesse *ranking*, com 1891 ocorrências. No jornal FSP ocorre 57 vezes; no jornal ZH, 19; no JM, 26; e no DG, 423.

- O pronome **ele** aparece 221 no CE contra 2886 ocorrências no jornal FSP, 2068 no DG, 1698 no JM e 1518 no jornal ZH.

Diante disso, cremos que esta métrica reflete o que ocorre em nossos *corpora* em relação à incidência de pronomes.

8.9 (#) PROPORÇÃO DE ORAÇÕES RELATIVAS EM RELAÇÃO À QUANTIDADE DE ORAÇÕES DO TEXTO.

Esta métrica mede a quantidade de orações relativas em comparação com as demais orações do texto. Nos Estudos-Pilotos 3 e 4, verificamos vários constituintes frasais expandidos por orações relativas. Diante disso, considerando que esse tipo de oração, na literatura sobre letramento, leitura e compreensão, é considerado um fator a dificultar a compreensão leitora por demandar maior carga de trabalho, julgamos importante verificar como as orações relativas se apresentam na totalidade do CE e dos *corpora* de contraste.

A Tabela 8.10 apresenta os dados relativos a esta métrica. Quanto mais alto o índice, maior a proporção de orações relativas presentes no texto e, conseqüentemente, mais complexo tende a ser este texto.

Tabela 8.10 – Proporção de Orações Relativas em relação à Quantidade de Orações do Texto

CORPUS ANALISADO	JM	FSP	DG	CE	ZH
ÍNDICE	0,08171	0,0915	0,0921	0,0945	0,1143
ORDEM DE DIFICULDADE	1 (mais fácil)	2	3	4	5 (mais difícil)

Pelos dados da Tabela 8.10, o texto mais difícil quanto ao emprego de orações relativas foi o do jornal ZH, seguidos do CE. O resultado do CE, em princípio, confere

com os achados dos Estudos-Pilotos 3 e 4: vários constituintes frasais expandidos por orações relativas. O jornal ZH não foi alvo dessas análises, por isso não podemos fazer essa mesma comparação.

O DG, em relação à incidência de pronomes, foi o segundo mais difícil, seguido do jornal ZH, em terceiro. E aqui, quanto às orações relativas, ficou como o terceiro mais difícil, portanto, mais fácil que o jornal ZH. O JM ficou como o mais fácil quanto à incidência de orações relativas e o segundo mais fácil em relação à incidência de pronomes em geral.

O comportamento do DG e do JM, mesmo alterando em uma posição no *ranking* das duas métricas, manteve uma certa harmonia. É preciso ponderar que, na métrica dos pronomes, há muitos outros tipos além dos relativos e que as comparações com os estudos-pilotos realizados é uma comparação entre resultado de amostras com resultados da totalidade. Ainda que, em muitas situações, os resultados das amostras tenham sido equivalentes ao da totalidade, podem ocorrer variações.

Além disso, muitas das palavras morfologicamente marcadas como pronomes relativos podem também ter outras classificações e atuarem de modo diverso, como é o caso de **quanto**. A seguir trazemos alguns exemplos de nossos *corpora*.

- (8.12) É promissor que **tanto** a administração estadual **quanto** a da Capital estejam compromissadas com as mudanças encaminhadas ao Congresso. (ZH)
- (8.13) Tenho vários ases na manga, quero voltar com um quadro **tão** poderoso **quanto** este. (DG)
- (8.14) Além do mau exemplo do domingo, Bolsonaro tem insistido em que o coronavírus não é **tão** perigoso **quanto** parece, ignorando aspectos epidemiológicos que não têm a ver com a taxa de letalidade do patógeno. (FSP)
- (8.15) "É muito bom conhecer esse mundo da maternidade, **tanto** a parte boa **quanto** a parte ruim", afirmou, enquanto o amamentava. (JM)
- (8.16) Assim, cabe a parte contrária, ou seja, a Empresa Demandada no processo em questão, também produzir **tantas** provas **quanto** sejam necessárias, visando destruir a prova testemunhal produzida pela parte autora, haja vista o princípio do Contraditório. (CE)

Segundo o Houaiss, nesses casos **quanto** é considerado um advérbio empregado como correlativo de *tão* e *tanto*: “tanto estuda q. se diverte”.

Na ferramenta Sketch Engine, as ocorrências de **quanto** nos exemplos acima estão listadas entre as situações em que **quanto** foi considerado **um pronome**. Apesar de os dados da métrica **Incidência de orações relativas** serem extraídos pelo sistema NILC-

Metrix, pode haver dificuldade na delimitação dessas ocorrências. Desse modo, os índices podem não revelar com exatidão o comportamento dos *corpora* nesse quesito.

A Tabela 8.11 apresenta os dados referentes à métrica Proporção de orações relativas em relação à quantidade de orações do texto em comparação com dados referentes à Incidência de Pronomes e de Pronomes Relativos nos *corpora* analisados.

Tabela 8.11 – Comparação entre a Proporção de Orações Relativas em relação à Quantidade de Orações do Texto, a Incidência de Pronomes em geral e a Incidência de Pronomes Relativos

<i>Corpus</i> analisado	JM	FSP	DG	CE	ZH
Índice de orações relativas	0,08171	0,0915	0,0921	0,0945	0,1143
Ordem de dificuldade	1 (mais fácil)	2	3	4	5 (mais difícil)
Índice de pronomes	0,0413	0,05916	0,0468	0,0406	0,0457
Ordem de dificuldade	2	5 (mais difícil)	4	1 (mais fácil)	3
Índice de pronomes relativos	0,4504	0,4548	0,4017	0,5546	0,4924
Ordem de dificuldade	2	3	1 (mais fácil)	5 (mais difícil)	4

Os dados da Proporção de Orações Relativas e da Incidência de Pronomes foram extraídos das Tabelas 8.9 e 8.10. Os índices da Incidência de Pronomes Relativos foram calculados pela divisão do total de pronomes relativos pelo total de pronomes em cada *corpus*, com base nos dados extraídos pela ferramenta *Sketch Engine*. Considerando que os pronomes relativos sempre iniciam uma oração relativa, julgamos pertinente a comparação apresentada na Tabela 8.11.

Se atentarmos para as linhas que estabelecem a ordem de dificuldade de cada *corpus* em relação a cada índice apresentado, veremos que não obedecem à mesma ordenação.

DG: ficou em terceiro lugar em relação à incidência de orações relativas, mas como o mais fácil em relação à incidência de pronomes relativos. Em relação aos pronomes totais, ficou como o segundo mais difícil. A diferença entre a incidência de

orações relativas e a de pronomes relativos não deveria ser tão grande, uma vez que os pronomes relativos sempre iniciam uma oração relativa.

FSP: ficou como segundo mais fácil em relação à incidência de orações relativas, ficou como mais difícil em relação à totalidade dos pronomes e terceiro colocado quanto a apenas pronomes relativos. Aqui também se verifica diferença nas colocações quanto às orações relativas e pronomes relativos, apesar de ser em apenas 1 colocação.

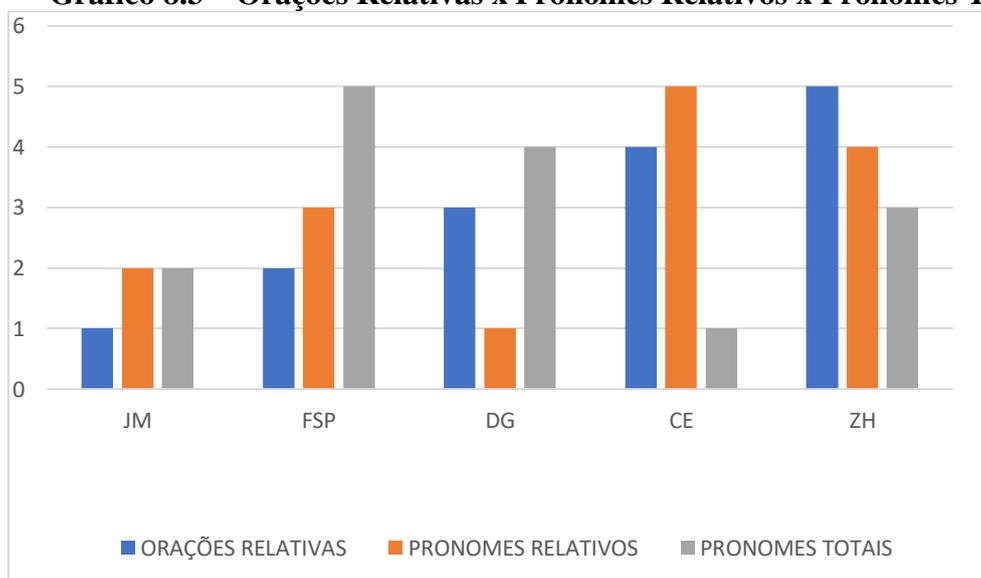
JM: de mais fácil quanto à incidência de orações relativas, ficou como o segundo mais fácil tanto em relação aos pronomes totais como em relação aos pronomes relativos.

CE: ficou como o segundo mais difícil em relação à incidência de orações relativas, o segundo mais difícil em relação aos pronomes relativos e o mais fácil em relação aos pronomes.

ZH: ficou como o mais difícil quanto às orações relativas, e o segundo mais difícil quanto aos pronomes relativos. Em relação aos pronomes em geral, ficou na terceira colocação.

Para melhor visualização desse *ranking*, abaixo colocamos o Gráfico 8.3.

Gráfico 8.3 – Orações Relativas x Pronomes Relativos x Pronomes Totais



O Gráfico 8.3 segue a lógica dos gráficos anteriores: o eixo vertical à esquerda representa os níveis de dificuldade, sendo o 1 o mais fácil e o 5 o mais difícil. Pelo Gráfico 8.3, percebe-se um comportamento inversamente proporcional entre o jornal ZH e o CE no que se refere às orações relativas e aos pronomes relativos.

Relembramos aqui que o DG foi o *corpus* que mais apresentou constituintes frasais pronominais e que algumas palavras marcadas como pronomes podem não se comportar como tal nos textos, como exemplificamos com a palavra **quanto**. Diante disso, cremos que os índices que mais se aproximam da realidade dos *corpora* de nosso estudo quanto à Incidência de Orações Relativas são os referentes à incidência de pronomes relativos. Se assim for considerado, o texto mais complexo será o CE, seguido do jornal ZH. E o mais fácil será o DG, seguido do JM.

8.10 INCIDÊNCIA DE CONECTIVOS TOTAL

Esta métrica mede a proporção de conectivos em relação às demais palavras do texto. Na fundamentação dos Estudos-Pilotas 3 e 4, verificamos que, entre as características analisadas para verificar a complexidade sentencial, estão os processos de coordenação e de subordinação, intimamente relacionados à presença de conectivos.

Além disso, conforme dados do Estudo-Piloto 1, o CE apresentou um índice menor de conectivos em relação aos demais *corpora*. Segundo referimos naquele momento, a baixa incidência de conectivos pode estar relacionada à presença de muitas orações subordinadas reduzidas (muito comuns em textos jurídicos). E as orações reduzidas nem sempre são claras na relação que estabelecem, como ocorre frequentemente com o emprego do gerúndio, por exemplo. E isso pode ser um fator que dificulte a inteligibilidade do texto.

Por essas razões e para verificar como se comportam os *corpora* em sua totalidade, julgamos pertinente trazer os dados desta métrica às análises aqui apresentadas.

A Tabela 8.12 apresenta, então, os dados sobre a incidência de conectivos em nossos *corpora*.

Tabela 8.12 – Incidência de Conectivos Total

CORPUS ANALISADO	ZH	FSP	JM	DG	CE
ÍNDICE	0,0794	0,07921	0,0760	0,0684	0,0625
ORDEM DE DIFICULDADE	1(mais fácil)	2	3	4	5(mais difícil)

Quanto maior o valor desta métrica, que significa maior incidência de conectivos no texto, mais fácil é este texto. Assim, conforme vemos na Tabela 8.12, o texto mais fácil é do jornal ZH, e o mais difícil é o CE. O JM e o DG, jornais populares, ficaram mais difíceis que os jornais FSP e ZH.

O resultado do CE atual confere com os índices do Estudo-Piloto 1. Nesse estudo-piloto, o CE também foi o texto que apresentou menor incidência de conectivos.

Quanto aos Estudos-Pilotos 3 e 4, a maioria dos constituintes frasais oracionais presentes no CE apresentaram o conectivo expresso. E o índice menor do CE nesta métrica parece ser contraditório. Ocorre que os conectivos presentes nessas orações são os conhecidos como “conjunções integrantes”, e a métrica Incidência de Conectivos Total verifica não só a presença dessas conjunções integrantes, como das demais conjunções existentes na Língua Portuguesa. Além disso, o Manual do NILC-Matrix refere como uma limitação da métrica o fato de uma mesma palavra poder atuar como conectivo em uma situação e não em outra, o que poderia levar a uma certa imprecisão. É o caso da palavra **segundo**, nos seguintes exemplos extraídos do DG.

(8.17) Os outros estudantes já haviam saído da sala, **segundo** a ocorrência policial. (conjunção subordinativa)

(8.18) Em contrapartida, o vice de futebol, Giovanni Luigi, admitiu que o clube deve vender um atleta no **segundo** semestre. (numeral ordinal)

Diante disso, cremos que esta métrica, apesar das suas limitações, espelhe o que ocorre com o CE. Para uma análise mais acurada, se poderia verificar a quantidade de formas verbais nominais, o que daria um indício da quantidade de orações reduzidas. Ainda assim, seria necessário distinguir as situações em que essas formas verbais nominais estariam realmente introduzindo orações reduzidas ou fazendo parte de uma locução verbal.

Em relação aos outros *corpora*, chamou-nos a atenção o fato de o DG e o JM terem ficado mais complexos que os jornais ZH e FSP. No Estudo-Piloto 1, ficaram como mais fáceis. Como já referimos, naquela oportunidade trabalhamos com um conjunto pequeno de textos, em torno de 8 mil *tokens* cada *corpus*. A análise aqui se dá em *corpora* de aproximadamente 500 mil palavras. É possível haver variações como a que ocorre aqui.

8.11 (#) PROPORÇÃO DE ORAÇÕES SUBORDINADAS PELA QUANTIDADE DE ORAÇÕES DO TEXTO E PROPORÇÃO DE CONJUNÇÕES SUBORDINATIVAS EM RELAÇÃO A TODAS AS CONJUNÇÕES DO TEXTO

A primeira métrica avalia a proporção de orações subordinadas em relação às demais orações do texto. A segunda verifica a proporção de conjunções subordinativas em relação a todas as conjunções do texto.

Incluímos estas métricas em nosso estudo pelo fato de, nos Estudos-Pilotos 3 e 4, termos verificado, em vários exemplos trazidos, que as frases do CE se organizavam em períodos longos de uma frase só com muitas orações subordinadas em vários níveis de hierarquia. Como já referimos, na literatura sobre o letramento e compreensão leitora, esse tipo de organização frasal é outro complicador para a leitura.

A escolha dessas duas métricas pode parecer redundante, porém não é. Podem ocorrer orações subordinadas reduzidas, como vimos entre os constituintes frasais oracionais analisados nos Estudos-Pilotos 3 e 4. Desse modo, a análise dos índices relativos à proporção de conjunções subordinativas poderá ajudar a verificar se há alguma diferença significativa entre orações reduzidas e desenvolvidas. Nos Estudos-Pilotos 3 e 4, as desenvolvidas foram em maior número.

A Tabela 8.13 apresenta os índices relativos à Proporção de orações subordinadas pela quantidade de orações do texto e em relação à Proporção de Conjunções Subordinativas em relação a todas as conjunções do texto.

Tabela 8.13 – Proporção de Orações Subordinadas pela Quantidade de Orações do Texto e Proporção de Conjunções Subordinativas em relação a todas as Conjunções do Texto

CORPUS ANALISADO	JM	FSP	DG	ZH	CE
PROPORÇÃO DE ORAÇÕES SUBORDINADAS	0,3238	0,3295	0,3345	0,4304	0,4442
ORDEM DE DIFICULDADE	1(mais fácil)	2	3	4	5(mais difícil)
PROPORÇÃO DE CONJUNÇÕES SUBORDINATIVAS	0,1849	0,2553	0,1913	0,3037	0,3603
ORDEM DE DIFICULDADE	1(mais fácil)	3	2	4	5(mais difícil)

Pelos dados da Tabela 8.13, o CE se manteve como o *corpus* mais difícil em relação às orações subordinadas e às conjunções subordinativas. Tal resultado confere com os achados dos Estudos-Pilotos 3 e 4, em que as amostras de CE apresentaram mais constituintes frasais oracionais e, entre esses, mais orações desenvolvidas.

O JM e o jornal ZH também mantiveram sua colocação. A ZH em segundo mais difícil, e o JM como o mais fácil.

No Estudo-Piloto 4, o JM apresentou menos constituintes frasais oracionais que as sentenças, mas mais que o DG. De todo modo, as orações que representam os constituintes frasais analisados nesse Estudo-Piloto 4 são as chamadas “substantivas” e não contemplam todo o conjunto de orações subordinadas. Existem ainda as orações subordinadas adverbiais e adjetivas. cremos, então, que esse resultado é viável.

Chama a atenção o comportamento do DG e do jornal FSP. O DG, de terceira posição em relação às orações subordinadas, passa a segundo em relação à presença de conjunções subordinativas. E o jornal FSP, de terceiro em relação às orações subordinadas, passa a segundo quanto às conjunções subordinativas.

Quanto a essa situação, inicialmente se pode pensar na presença de orações desenvolvidas e reduzidas. O fato de o DG ser o terceiro em relação à presença de orações e segundo em relação à presença de conjunções subordinativas pode significar que, entre as orações subordinadas que apresenta, a maioria são reduzidas. Entre os sujeitos oracionais encontrados no Estudo-Piloto 4, foram 9 ocorrências de orações reduzidas contra 3 desenvolvidas. Entre os complementos verbais oracionais, foram 52 desenvolvidas contra 37 reduzidas. Nesse último caso, houve mais desenvolvidas que reduzidas. Porém, é preciso lembrar que as orações analisadas no Estudo-Piloto 4 não contemplam, como já dissemos, o universo de todas as orações subordinadas e, conseqüentemente, de todas as conjunções subordinativas.

O processo inverso ao do DG parece ocorrer com o jornal FSP, que não foi analisado no Estudo-Piloto 4. O fato de apresentar menos orações subordinadas e mais conjunções subordinativas que o DG pode significar a presença de menos orações reduzidas e mais orações desenvolvidas.

Os dados por estas métricas aqui expostos confirmam as nossas observações quanto à organização sintática do CE com a presença de muitos períodos longos de uma frase só e vários níveis de subordinação. Ele se manteve como o mais difícil nas duas métricas aqui analisadas.

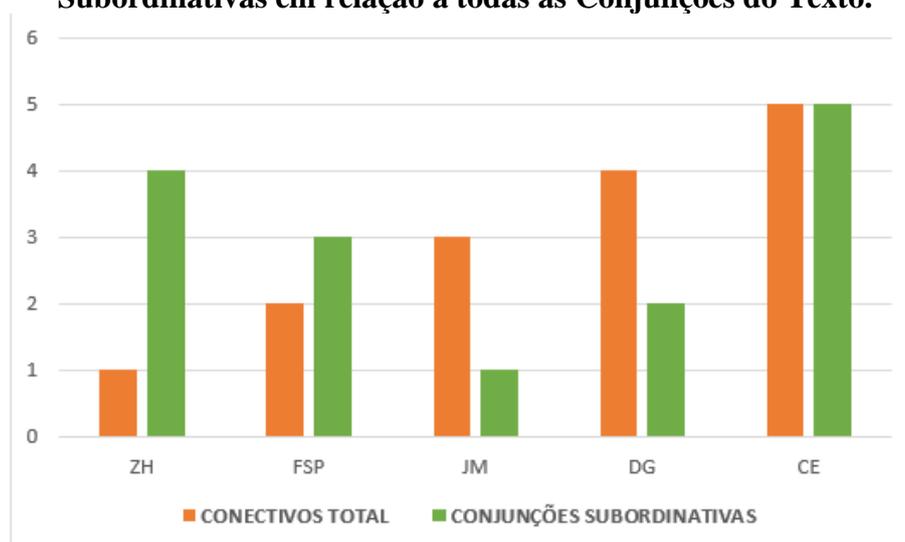
Como as conjunções subordinativas são um tipo de conectivo, outra relação que julgamos importante estabelecer neste momento é quanto à Incidência de Conectivos Total e a Proporção de Conjunções Subordinativas em relação ao total de conjunções do texto. A Tabela 8.14 traz esses dados.

Tabela 8.14 – Incidência de Conectivos Total x Proporção de Conjunções Subordinativas em relação a todas as Conjunções do Texto

CORPUS ANALISADO	ZH	FSP	JM	DG	CE
CONNECTIVOS TOTAL	0,0794	0,07921	0,0760	0,0684	0,0625
ORDEM DE DIFICULDADE	1(mais fácil)	2	3	4	5(mais difícil)
CONJUNÇÕES SUBORDINATIVAS	0,3037	0,2553	0,1849	0,1913	0,3603
ORDEM DE DIFICULDADE	4	3	1(mais fácil)	2	5(mais difícil)

Para facilitar a compreensão da Tabela 8.14 acima, apresentamos o Gráfico 8.4, abaixo:

Gráfico 8.4 – Incidência de Conectivos Total x Proporção de Conjunções Subordinativas em relação a todas as Conjunções do Texto.



Considerando o eixo vertical do Gráfico 8.4, que indica o nível de dificuldade dos textos (1 – mais fácil; e 5 – mais difícil), é possível verificar que o comportamento dos jornais em relação a essas métricas oscila bastante:

ZH: de mais fácil em relação aos conectivos totais, passou para o segundo mais difícil em relação às conjunções subordinativas.

FSP: mudou apenas a 1 posição.

JM: de terceiro mais fácil em relação aos conectivos total passou para o primeiro mais fácil em relação às conjunções subordinativas.

DG: de segundo mais difícil em relação aos conectivos passou para o segundo mais fácil em relação às conjunções subordinativas.

Essa oscilação é possível porque as conjunções subordinativas são uma espécie de conectivos, mas não contemplam a sua totalidade. A maior incidência de conectivos em um texto é um fator facilitador, como já referimos, pois as ideias ficam mais visíveis e, portanto, mais claras. Entretanto, a maior incidência de conjunções subordinativas não é considerada um fator facilitador, pois introduzem orações subordinadas, que, por sua vez, indicam estruturas mais complexas.

No caso do jornal ZH, por exemplo, houve maior presença de conectivos (índice considerado facilitador do texto). Dentro deste conjunto, houve uma maior incidência de conjunções subordinativas, índice considerado dificultador.

Diante disso, em termos de complexidade, os jornais FSP e ZH, por apresentarem mais conjunções subordinativas, são mais difíceis que os jornais populares DG e JM, que apresentam mais conectivos e menos conjunções subordinativas.

O CE é que se mantém o mais difícil de todos em ambas as métricas. Este comportamento reforça o nível de dificuldade que esse tipo de texto oferece em relação às conexões que são estabelecidas entre as ideias do texto. Além disso, confirma as observações que até aqui temos feito em relação à complexidade dos períodos longos e com vários níveis de subordinação presentes nos textos do CE.

8.12 (#) PROPORÇÃO DE SUJEITOS POSPOSTOS EM RELAÇÃO A TODOS OS SUJEITOS DO TEXTO

Esta métrica será analisada em conjunto com a métrica a seguir, relacionada com a proporção de orações que não estão no formato SVO (sujeito-verbo-objeto) em relação a todas as orações do texto.

8.13 (#) PROPORÇÃO DE ORAÇÕES QUE NÃO ESTÃO NO FORMATO SVO (SUJEITO-VERBO OBJETO) EM RELAÇÃO A TODAS AS ORAÇÕES DO TEXTO.

Estas métricas medem a proporção de sujeitos pospostos ao verbo em relação a todos os sujeitos do texto e também a proporção de orações que não estão no formato SVO (sujeito-verbo-objeto) em relação a todas as orações do texto.

Escolhemos estas métricas porque a alteração da ordem canônica SVO é fator considerado dificultador para a compreensão leitora. No Estudo-Piloto 3, os sujeitos pós-verbais representaram 16% do total de sujeitos da amostra. No Estudo-Piloto 4, as sentenças mantiveram um índice aproximado ao do Estudo-Piloto 3, com 16,9% de sujeitos pós-verbais. Essa ocorrência representou mais que o dobro das ocorrências desse tipo de sujeito no DG, com 6,8%, e no JM, com 6,4%. Diante desses resultados, os textos das sentenças se mostraram mais complexos que os dos jornais.

Ressalvamos que nem todas as orações que não estão no formato SVO (sujeito-verbo-objeto) apresentam sujeito posposto, como é o caso, por exemplo, das orações sem sujeito ou das que apresentam sujeito elíptico. Entretanto, todos os sujeitos pospostos fazem parte de uma oração que não está no formato SVO. Por essa razão, decidimos analisar os índices das duas métricas conjuntamente e verificar em que medida o comportamento dos *corpora* se mantém em relação a elas.

A Tabela 8.15 apresenta os dados relativos a estas métricas. Quanto maior o índice da métrica, maior é a complexidade do texto, porque há maior incidência de sujeitos pospostos e de orações que não estão no formato SVO.

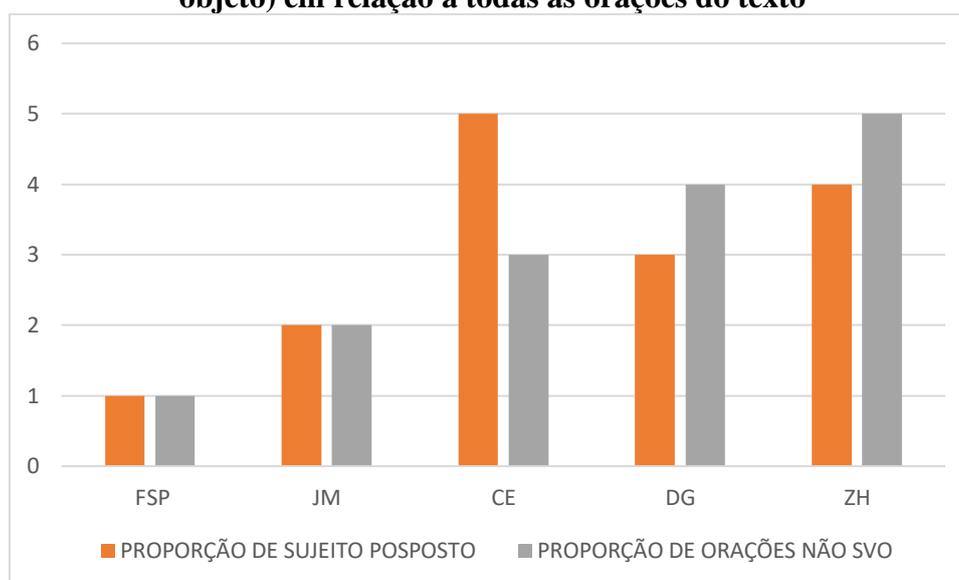
Tabela 8.15 – Proporção de Sujeitos Pospostos em relação a todos os Sujeitos do Texto e Proporção de Orações que não estão no formato SVO (sujeito-verbo-objeto) em relação a todas as Orações do Texto

CORPUS ANALISADO	FSP	JM	CE	DG	ZH
PROPORÇÃO DE SUJEITO POSPOSTO	0,0657	0,0774	0,2031	0,0807	0,0987
ORDEM DE DIFICULDADE	1 (mais fácil)	2	5 (mais difícil)	3	4

PROPORÇÃO DE ORAÇÕES NÃO SVO	0,2192	0,2772	0,2979	0,3030	0,3136
ORDEM DE DIFICULDADE	1 (mais fácil)	2	3	4	5 (mais difícil)

Os dados da Tabela 8.15 acima estão expostos no Gráfico 8.5, abaixo.

Gráfico 8.5 – Proporção de sujeitos pospostos em relação a todos os sujeitos do texto e Proporção de orações que não estão no formato SVO (sujeito-verbo-objeto) em relação a todas as orações do texto



Com base nos dados da Tabela 8.15 e no Gráfico 8.5, é possível verificar inicialmente que os jornais FSP e JM mantiveram o mesmo *ranking* nas duas medidas, ficando o jornal FSP como mais fácil, e o JM como o segundo mais fácil.

O fato de o jornal FSP ter ficado como o mais fácil nas duas medidas chama a atenção, por ser um jornal de referência e voltado a um público em princípio mais escolarizado que o público dos jornais populares. E o JM confirmou os dados do Estudo-Piloto 4, em que ficou com menos sujeitos pospostos que o DG e a amostra de sentenças.

O jornal ZH ficou com mais difícil em relação às orações que não estão no formato SVO e o segundo mais difícil em relação aos sujeitos pospostos. Tendo em vista o tipo de jornal que é, esse resultado não surpreende. Entretanto, surpreende um pouco o fato de ter ficado mais complexo que as sentenças quanto às orações que não estão no formato SVO. O DG, por sua vez, ficou como o segundo mais difícil em relação às orações que não estão no formato SVO e o terceiro mais difícil em relação aos sujeitos pospostos. Esse resultado do DG, apesar de confirmar os dados do Estudo-Piloto 4, em que ficou

com mais sujeitos pospostos que o JM, também causa perplexidade, por ser um jornal popular.

O CE ficou como o mais difícil em relação aos sujeitos pospostos, confirmando os achados do Estudo-Piloto 4, em que apresentou mais que o dobro de sujeitos pospostos que os textos dos jornais. Em relação às orações que não estão no formato SVO, ficou com o terceiro lugar.

De todo modo, os dados destas métricas confirmam os resultados dos Estudos-Pilotos 3 e 4 quanto à posição do sujeito na oração e confirmam a maior complexidade do CE em relação a esse quesito. Em relação às orações que não estão no formato SVO, como abrangem outras formas de organização da oração, é possível haver essa variação nos resultados.

8.14 QUANTIDADE MÉDIA DE REFERENTES QUE SE REPETEM NOS PARES DE SENTENÇAS ADJACENTES DO TEXTO

Esta métrica verifica as palavras que se repetem entre duas sentenças adjacentes, ou seja, que estejam lado a lado. No Manual do NILC-Metrix, consta o seguinte exemplo para ilustrar como essa métrica funciona.

Exemplo do Manual do NILC-Metrix

*As crianças aprendem muito rápido. Pesquisas mostram que até os três anos de vida, o **desenvolvimento (1)** do cérebro ocorre num ritmo bem acelerado. O que os **pais (2)** fazem no dia-a-dia, como ler, cantar e demonstrar carinho, é crucial para o **desenvolvimento (1)** saudável da criança. Mas de acordo com certo estudo, apenas cerca da metade dos **pais (2)** com crianças entre dois e oito anos lê diariamente para elas. Você talvez se pergunte: 'Será que ler para o meu filho realmente faz diferença?'*

Na segunda frase do trecho acima, temos a repetição da palavra **desenvolvimento**, que apareceu na frase anterior. E, na terceira frase, temos a repetição da palavra **pais**, que apareceu na segunda oração.

No Estudo-Piloto 3 observamos algumas situações de sujeitos pronominais ou de construções com gerúndio em que a busca pelo referente não era tão clara. Os dados das métricas relativas aos referentes que se repetem nos pares de sentenças do texto, adjacentes ou não, podem ser utilizados para verificar em que medida esse fenômeno ocorre na totalidade de nosso CE e nos *corpora* de contraste.

A Tabela 8.16 apresenta os índices de nossos *corpora* em relação a essa métrica. Quanto maior o índice, maior é a repetição de palavras e, conseqüentemente, mais fácil é o texto.

Tabela 8.16 – Quantidade Média de Referentes que se repetem nos Pares de Sentenças Adjacentes do Texto

CORPUS ANALISADO	FSP	CE	JM	ZH	DG
ÍNDICE	0,6587	0,6556	0,5736	0,5080	0,3152
ORDEM DE DIFICULDADE	1 (mais fácil)	2	3	4	5(mais difícil)

Conforme se vê na Tabela 8.16, o DG ficou como o texto mais complexo. Ou seja, o DG é o texto que apresenta menor repetição de palavras. Esse dado é confirmado pela maior variação lexical que apresenta em relação aos outros *corpora*, conforme se vê na Tabela 8.1.

Causa estranheza o fato de o jornal FSP ficar como o texto mais fácil nesse quesito, visto que sua variação lexical foi superior à do CE. A variação lexical do jornal FSP foi de 5,57, e a do CE, de 2, 71.

O CE ficou como o segundo texto mais fácil. Este resultado se aproxima do que se deu no Estudo-Piloto 1. Quanto à métrica Sobreposição de Argumentos Adjacentes, do Coh-Metrix Port 2.0, que também mede as palavras que se repetem entre sentenças adjacentes do texto, o conjunto de sentenças ficou como o texto mais fácil. Além disso, segundo a Tabela 8.1, foi o *corpus* com menor variação lexical – 2,71% -, o que confirma o resultado apontado nesta métrica, com a ressalva para o resultado relativo ao jornal FSP.

O exemplo 8.16 a seguir comprova essa repetição de referentes nas sentenças.

Exemplo (8.16)

Não assiste razão ao **autor**.

O **autor** não comprovou que adimpliu o prêmio do seguro contratado, conforme documentos apresentados pelo próprio **autor** às fls.18 e 21-23, onde se verifica que o seguro contratado tinha como prêmio a quantia de R\$569,69, a ser paga em duas parcelas no valor de R\$284,87, com vencimento em 08.03.2017 e 03.04.2017. O **autor** apresentou comprovante de pagamento no valor de R\$361,74, pago em 23.06.2017, e nada mais.

Neste exemplo, a palavra **autor** aparece nas três frases do texto. Na segunda, inclusive, aparece duas vezes.

Pelos índices da TTR na Tabela 8.1, o jornal ZH deveria ficar em terceiro lugar, o JM em quarto nesta métrica. Mas aqui ocorre o inverso: o JM fica em terceiro, e o jornal ZH em quarto.

Agora vejamos como se comportam os *corpora* em relação à repetição das palavras no texto como um todo, segundo a métrica a seguir.

8.15 QUANTIDADE MÉDIA DE REFERENTES QUE SE REPETEM NOS PARES DE SENTENÇAS DO TEXTO

Esta métrica verifica as palavras que se repetem entre as frases do texto. A diferença em relação à métrica anterior é que, por esta métrica, as frases não precisam ser adjacentes. Para explicar o funcionamento desta métrica, o Manual do NILC-Metrix traz o mesmo exemplo dado na anterior, que repetimos aqui.

Exemplo do Manual do NILC-Metrix

As **crianças (1)** aprendem muito rápido. Pesquisas mostram que até os três **anos (2)** de vida, o **desenvolvimento (3)** do cérebro ocorre num ritmo bem acelerado. O que os **pais (4)** fazem no dia-a-dia, como ler, cantar e demonstrar carinho, é crucial para o **desenvolvimento (3)** saudável da criança. Mas de acordo com certo estudo, apenas cerca da metade dos **pais (4)** com **crianças (1)** entre dois e oito **anos (2)** lê diariamente para elas. Você talvez se pergunte: ‘Será que ler para o meu filho realmente faz diferença?’

No trecho acima, há 4 palavras que se repetem:

crianças (1), nas primeira e quarta frases;

anos (2), nas segunda e quarta frases;

desenvolvimento (3), nas segunda e terceira frases; e

pais (4), na terceira e quarta frases.

A Tabela 8.17 apresenta os índices de nossos *corpora* em relação a essa métrica. Quanto maior o índice, maior é a repetição de palavras e, conseqüentemente, mais fácil é o texto.

Tabela 8.17 – Quantidade Média de Referentes que se repetem nos Pares de Sentenças do Texto

<i>CORPUS ANALISADO</i>	JM	ZH	FSP	CE	DG
ÍNDICE	0,5048	0,4351	0,3592	0,3137	0,2618
ORDEM DE DIFICULDADE	1(mais fácil)	2	3	4	5(mais difícil)

O DG fica como o mais complexo novamente. Conforme já explicamos na métrica anterior, tal resultado é o esperado. Chama a atenção, porém, o fato de o CE ficar como o segundo mais difícil quanto aos referentes que se repetem entre as sentenças do texto como um todo. Considerando que o CE é o texto que apresenta menor variação lexical, o esperado seria que, nessas duas métricas, ele ficasse entre os mais fáceis. Essa repetição de referentes é exemplificada com no exemplo 8.17.

Exemplo (8.17)

É o caso de improcedência da **demanda**.

Quanto a **preliminar** de incompetência do JEC:

Em sua contestação, postulam os Requeridos Imobiliária XXXX XXXXX XXXXX o reconhecimento da incompetência do Juizado Especial Civil para processar a presente **demanda**, [...].

Tenho que a **preliminar** não merece ser acolhida, eis que conforme consulta ao sistema Themis, a referida ação foi distribuída em 07/07/2017 sendo que a presente **demanda** foi distribuída em data anterior, qual seja, 05/05/2017.

Quanto a **preliminar** de Ilegitimidade Passiva:

Em sua contestação, aduz a **Requerida** XXXXXXXX que é parte ilegítima para figurar no polo passivo da **demanda**, uma vez que apenas intermediou o contrato de locação entre a autora e o proprietário do imóvel, o **requerido** XXXXXXXX.

Tenho que a **preliminar** merece ser acolhida, eis que conforme prova carreada aos autos a **Requerida** XXXXXXXX Imóveis, atuou apenas como mandatária do **Requerido** XXXXXXXX, na intermediação do contrato de locação do imóvel.

Por essas razões, a extinção do feito, sem o julgamento do mérito, no que se refere à **Requerida** XXXXXXXX é a medida que se impõe.

Quanto ao mérito:

Quanto ao mérito, a **demanda** é improcedente [...]

No trecho acima:

- **demanda** aparece nas primeira, segunda, terceira, quinta e nona frases.
- **preliminar** aparece nas segunda, quarta, quinta e sétima frases.
- **requerida** aparece nas sexta, sétima e oitava frases.
- **requerido** aparece nas sexta e sétima frases.

Além disso, no Estudo-Piloto 4, o CE ficou com menos constituintes frasais pronominais e com menor incidência de pronomes. Esses índices levam a uma maior repetição das palavras, o que confere com a menor variação lexical que este texto apresenta.

Uma possível razão para essa menor repetição de argumentos entre as sentenças dos textos pode estar no emprego de sujeitos elípticos, que foram os de maior ocorrência na Amostra 1 e na Amostra 2 das sentenças. Abaixo colocamos um exemplo desse tipo de situação:

Exemplo (8.18)

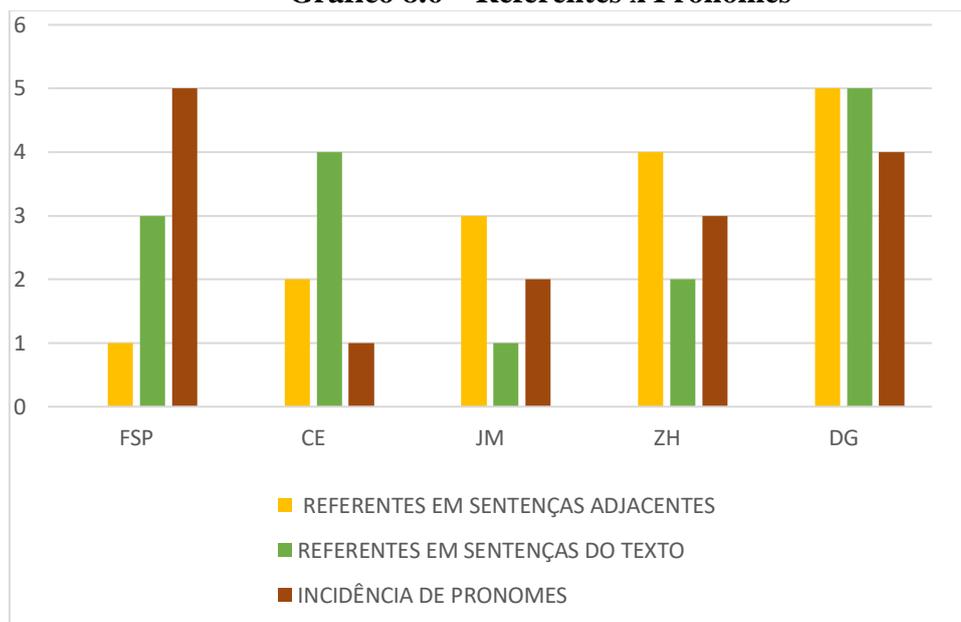
A parte requerida apresentou contestação escrita, onde **alegou** preliminarmente ausência de interesse processual do autor, já que o pedido do autor de revisão tarifária foi deferido administrativamente em outubro de 2016. **Ainda**, em preliminar, **alega** prescrição, tendo em vista que a presente ação foi protocolada em 19/04/2017 e a prescrição no presente caso é de 10 anos, não há que se falar em faturas anteriores a 19.04.2007.

No mérito, **descreve** a situação fática envolvendo o imóvel e também **faz** uma explanação sobre os tipos de tarifa e serviços realizados pela demandada. **Alega** ainda, que os atos praticados foram legais, e que assim que tomou conhecimento dos fatos, alterou a categoria tarifária do imóvel do autor.

Por fim, **se insurge** contra a devolução em dobro e inversão do ônus da prova, e a observância do artigo 100 da CF. **Juntou** documentos (folhas 258-412), **requerendo** a improcedência de todos os pedidos do autor.

Para melhor visualização, sublinhamos as palavras iniciais de cada frase do trecho acima. O texto inicia com o sujeito exposto **a parte requerida**. Depois, na continuidade do texto, colocamos em negrito todos os verbos que se referem a este mesmo sujeito. Em 5 frases, este mesmo sujeito é retomado 7 vezes por meio de verbos. Em todas essas situações, poderia haver a repetição expressa do sujeito ou o emprego de algum pronome que o substituísse, mas o redator optou pela elipse, pois o contexto permite compreender a quem esses verbos se referem.

Abaixo colocamos o Gráfico 8.6, que apresenta o comportamento de nossos *corpora* em relação à incidência de pronomes e às métricas relativas à repetição de argumentos. Julgamos pertinente fazer essa correlação porque o emprego de pronomes é um recurso intimamente relacionado com a repetição ou não de argumentos.

Gráfico 8.6 – Referentes x Pronomes

O eixo vertical do Gráfico 8.6 segue a lógica dos gráficos anteriores, indicando o nível de dificuldade dos *corpora* – 1 (mais fácil) e 5 (mais difícil). É interessante observar, pelo Gráfico 8.6, o comportamento do jornal FSP, que, em relação à repetição de argumentos, segue os mesmos passos que o CE. Quanto à repetição de argumentos em sentenças adjacentes, ficou como o mais fácil, como se vê pela coluna amarela menor em relação às demais. Quanto à repetição entre os argumentos nas sentenças do texto como um todo, ficou como o terceiro mais difícil (coluna verde no 3). Porém, ficou como o texto mais difícil em relação à incidência de pronomes, ou seja, foi o texto que apresentou maior frequência de pronomes (coluna marrom no 5). Em princípio, quanto mais pronomes, menor é a repetição das palavras, o que é coerente com o resultado apresentado aqui.

O JM ficou como o texto mais fácil nesta métrica, e o jornal ZH como o segundo mais fácil (ver as colunas verdes). Mudaram duas posições em relação à métrica anterior, dos argumentos de sentenças adjacentes (ver as colunas amarelas). Em relação à incidência de pronomes, mudaram apenas 1 posição (colunas marrons): o JM ficou como o segundo mais fácil, e o jornal ZH como o terceiro mais fácil. Ainda que tenham mudado de posição no *ranking*, esses *corpora* mantiveram um comportamento mais ou menos uniforme e coerente em relação a essas métricas.

O DG, como se vê, foi o que apresentou comportamento mais estável em relação a todas as métricas. Ficou como o texto com menor repetição de palavras e o segundo que mais empregou os pronomes, o que é coerente.

8.16 PROPORÇÃO DE PALAVRAS QUE DENOTAM NEGAÇÃO EM RELAÇÃO À QUANTIDADE DE PALAVRAS DO TEXTO

Esta métrica mede a proporção de palavras que denotam negação em relação à quantidade de palavras do texto. Conforme Pinker (2018, 2017), toda negação requer trabalho mental, porque nossa mente sempre procura a verdade. Assim, entender a frase negativa *O rei não está morto* é mais difícil do que entender a frase afirmativa *O rei está vivo*.

Segundo essa lógica, primeiro processamos a afirmação para depois processar a negação. É por isso que, entre as recomendações dos manuais de redação, está a de preferir frases afirmativas a negativas. Assim, no caso desta métrica, quanto maior o índice, mais complexo é o texto, pois apresenta maior incidência de palavras que denotam negação.

As palavras computadas por esta métrica são *não*, *nem*, *nunca*, *jamais* e *tampouco*. A negação não se dá apenas com o emprego dessas palavras, mas de outras com denotação negativa e também por meio de outros processos, como o emprego de prefixos negativos. Todavia, como no Estudo-Piloto 1, os jornais populares ficaram mais complexos que os demais em relação à negação, julgamos pertinente verificar se esse *ranking* se daria em relação à totalidade dos *corpora* de contraste em comparação com o CE.

A seguir, na Tabela 8.18, apresentamos os índices desta métrica.

Tabela 8.18 – Proporção de Palavras que Denotam Negação em relação à Quantidade de Palavras do Texto

CORPUS ANALISADO	JM	DG	ZH	FSP	CE
ÍNDICE	0,005333	0,006773	0,007504	0,009415	0,010573
ORDEM DE DIFICULDADE	1 (mais fácil)	2	3	4	5 (mais difícil)

Como se vê na Tabela 8.18, o *ranking* do Estudo-Piloto 1 se alterou: os jornais populares ficaram como mais fáceis, e o CE como o mais difícil nesse quesito. E aqui, tal qual fizemos na ocasião do Estudo-Piloto 1, revisamos a presença dessas palavras em todos os *corpora* analisados, e os dados da Tabela 8.18 se confirmaram.

Na totalidade dos *corpora*, os resultados seguem, então, o esperado: que os jornais populares sejam mais acessíveis quanto à presença da negação.

Após a exposição das métricas escolhidas com as considerações sobre os resultados apresentados em cada uma delas, para verificarmos como os *corpora* aqui analisados se comportaram em relação à totalidade dessas medidas, apresentamos a Tabela 8.19.

Tabela 8.19 – Média das Métricas Analisadas

GRUPO	MÉTRICA	CORPORA ANALISADOS				
		CE	CCA		CCB	
			ZH	FSP	DG	JM
LEXICAIS	Média de pal. concretas	4	3	5	1	2
	Densidade Semântica	3	2	1	5	4
	Idade de Aquisição da Palavra	5	3	4	1	2
	Palavras por frase	2	4	5	1	3
	Incidência de Adjetivos	5	3	4	1	2
LEITURABILIDADE	Índice Flesch	3	4	5	1	2
	Estatística de Honoré	1	3	2	5	4
SINTAXE	Incidência de Conectivos	5	1	2	4	3
	Distância na Árvore de Dependência	4	2	5	1	3
	Orações que não estão na ordem SVO	3	5	1	4	2
	Incidência de Sujeito Posposto	5	4	1	3	2
	Incidência de Pronomes	1	3	5	4	2
	Incidência de Conj. Subordinativas	5	4	3	2	1
	Incidência de Orações Relativas	4	5	2	3	1
	Incidência de Orações Subordinadas	5	4	2	3	1
	Arg. que se repetem em sentenças adjacentes	2	4	1	5	3
	Arg. que se repetem entre as sentenças do texto	4	2	3	5	1
	Palavras que denotam negação	5	3	4	2	1
	MÉDIA GERAL		3,7	3,3	3,1	2,8
ORDEM DE DIFICULDADE		5 (mais difícil)	4	3	2	1 (mais fácil)

Os dados relativos à Estatística de Honoré e ao Número de Palavras por Sentença na Tabela 8.19 foram computados com base nos dados da Tabela 8.1.

Como se vê na Tabela 8.19, em vez de as métricas virem acompanhadas por índices numéricos relativos a cada *corpus*, foram colocados números de 1 a 5. Essa numeração segue o procedimento que adotamos até este momento, ou seja, estabelece uma ordem de dificuldade da métrica conforme o número. Nessa escala, o número 1 representa o texto mais fácil, e o 5 o mais difícil da escala.

Na penúltima linha da tabela, temos a média geral de cada *corpus* em relação à totalidade das métricas analisadas. Com base nessa média, criamos a última linha da tabela, na qual apresentamos a numeração de 1 a 5. Nela, repetimos o procedimento adotado nas análises até aqui, para facilitar a leitura, de colocar ao lado do número 1 a indicação (mais fácil) e, ao lado do número 5, a indicação (mais difícil).

De acordo, então, com as últimas linhas da Tabela 8.19 o *corpus* que apresenta o texto mais difícil pela média das métricas analisadas é o CE, e o texto mais fácil é o do JM. O segundo mais fácil foi o DG, e o segundo mais difícil foi o jornal FSP. O jornal ZH ficou no meio termo, em terceiro lugar.

O resultado relativo aos jornais condiz com a finalidade de cada um deles. Os jornais populares ficaram mais fáceis, e os de referência mais complexos. O CE, por sua vez, ficou mais difícil, o que também é um resultado plausível, considerando as observações que até este momento fizemos acerca da sua sintaxe mais complexa e do seu léxico carregado de terminologia e palavras pouco comuns.

As métricas que utilizamos neste estudo analisaram os textos por vários critérios. Na primeira coluna da esquerda para a direita na Tabela 8.19, colocamos esses critérios (grupos) e as métricas que a eles correspondem, conforme esclarecemos à medida que as apresentamos.

O primeiro grupo corresponde às métricas que incluímos no grupo das Lexicais. A Tabela 8.20 apresenta essas métricas, e a média dos *corpora* em relação a elas apenas.

Tabela 8.20– Média das Métricas Lexicais

GRUPO	MÉTRICA	CORPORA ANALISADOS				
		CE	CCA		CCB	
			ZH	FSP	DG	JM
LEXICAIS	Média de pal. concretas	4	3	5	1	2
	Densidade Semântica	3	2	1	5	4
	Idade de Aquisição da Palavra	5	3	4	1	2
	Incidência de Adjetivos	5	3	4	1	2
MÉDIA		4,3	2,8	3,5	2	2,5
ORDEM DE DIFICULDADE		5 (mais difícil)	3	4	1 (mais fácil)	2

Assim como fizemos na Tabela 8.18, a última linha apresenta os números de 1 a 5, que indicam a ordem de dificuldade dos *corpora* em relação às métricas lexicais. Pelo que se vê, então, o CE ficou como o texto mais difícil em relação a aspectos lexicais dos *corpora*, e o DG ficou como o mais fácil. O resultado aqui segue o que ocorreu com a totalidade das métricas.

Vejamos agora, na Tabela 8.21, como fica o *ranking* em relação às métricas de leiturabilidade, que dizem respeito à extensão das frases e das palavras e a influência que exercem na compreensão leitora.

Tabela 8.21 – Média das Métricas de Leiturabilidade

GRUPO	MÉTRICA	CORPORA ANALISADOS				
		CE	CCA		CCB	
			ZH	FSP	DG	JM
LEITURABILIDADE	Índice Flesch	3	4	5	1	2
	Estatística de Honoré	1	3	2	5	4
	Palavras por frase	2	4	5	1	3
MÉDIA		2	3,7	4	2,3	3
ORDEM DE DIFICULDADE		1 (mais fácil)	4	5 (mais difícil)	2	3

Não incluímos aqui a TTR na Tabela 8.21 por duas razões. A primeira se deu porque seu cálculo neste trabalho foi com base nos dados relativos ao número de *tokens* e *types* extraídos da ferramenta AntConc, e não pelo sistema NILC-Metrix. E a segunda

se deu porque utilizamos a Estatística de Honoré, disponibilizada pelo Sistema NILC-Metrix. Essa métrica, como já explicamos, mede a relação entre os *tokens*, as palavras que aparecem uma única vez (os *hapax legomena*) e as palavras lexicais, ou palavras de conteúdo, do texto. Ou seja, seus índices trazem uma melhor ponderação da variação lexical.

Em relação a essas métricas de leiturabilidade, os textos mais difíceis foram os jornais de referência, FSP e ZH, que ficaram empatados, e o mais fácil foi o CE. Os jornais populares ficaram em segundo lugar.

Lembramos que o IF leva em conta o tamanho das palavras e das frases e, quando da análise do seu resultado, observamos que, muitas vezes, uma palavra pequena pode ser mais complexa que outra maior, como exemplificamos com **azo** e **responsabilidade**. Além disso, por apresentar muitos cabeçalhos de ementas em seu interior, o CE pode ter a contagem do número de palavras por frase também mascarado, afetando o resultado, como parece ser o caso.

Lembramos também que, em relação à Estatística de Honoré, tal como em relação às outras métricas relacionadas à variação lexical, o CE ficou sempre mais acessível. Assim ocorre, porque as sentenças são textos especializados com temáticas específicas, o que tende a diminuir a variação lexical.

Por tais razões, cremos que tais resultados em relação ao CE são plausíveis.

Outro grupo, o mais extenso, corresponde às métricas que medem a complexidade sintática dos textos. Algumas dessas métricas são enquadradas, no sistema NILC-Metrix, em outros grupos, que não o da complexidade sintática. Entretanto, nós as mantivemos juntas no grupo da Sintaxe, tendo em vista que atuam de modo interligado. É o caso, por exemplo, da incidência de conjunções subordinativas e de orações subordinadas, do emprego de argumentos que se repetem com o emprego de pronomes.

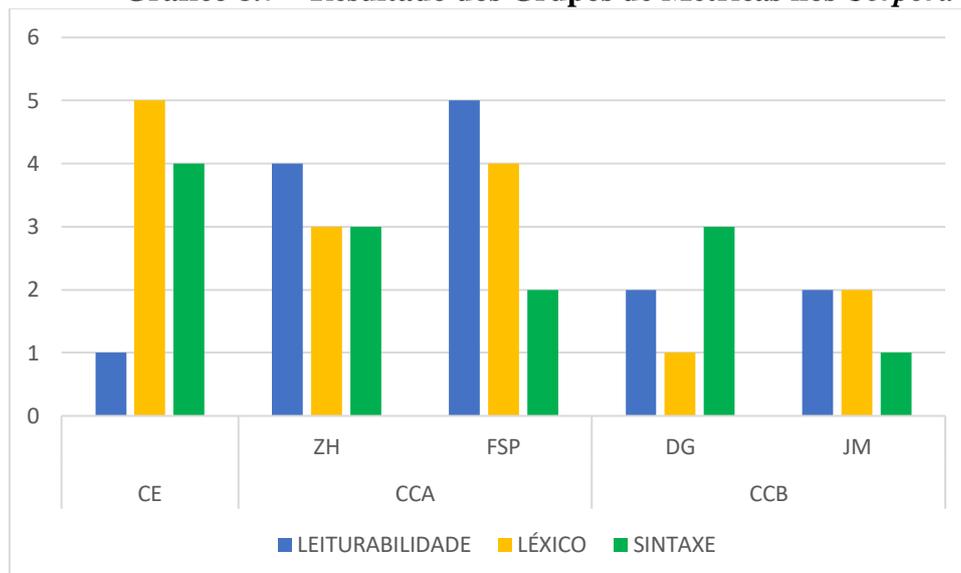
A Tabela 8.22 apresenta as médias deste grupo isoladamente.

Tabela 8.22 – Média das Métricas Sintáticas

GRUPO	MÉTRICA	CORPORA ANALISADOS				
		CE	CCA		CCB	
			ZH	FSP	DG	JM
SINTAXE	Incidência de Conectivos	5	1	2	4	3
	Distância na Árvore de Dependência	4	2	5	1	3
	Orações que não estão na ordem SVO	3	5	1	4	2
	Incidência de Sujeito Posposto	5	4	1	3	2
	Incidência de Pronomes	1	3	5	4	2
	Incidência de Conj. Subordinativas	5	4	3	2	1
	Incidência de Orações Relativas	4	5	2	3	1
	Incidência de Orações Subordinadas	5	4	2	3	1
	Arg. que se repetem em sentenças adjacentes	2	4	1	5	3
	Arg. que se repetem entre as sentenças do texto	4	2	3	5	1
	Palavras que denotam negação	5	3	4	2	1
MÉDIA		3,9	3,4	2,6	3,3	1,8
ORDEM DE DIFICULDADE		5 (mais difícil)	4	2	3	1 (mais fácil)

Como se vê na última linha da Tabela 8.22, o texto mais difícil do ponto de vista sintático é o do CE, e o mais fácil o do JM.

O Gráfico 8.7 abaixo apresenta o resultado dos *corpora* pelos grupos de métricas das Tabelas 8.20, 8.21 e 8.22.

Gráfico 8.7 – Resultado dos Grupos de Métricas nos *Corpora*

Tal como nos outros gráficos, o eixo vertical do Gráfico 8.7 apresenta a numeração por ordem de dificuldade dos *corpora*. O número 1 indica o texto mais fácil e o número 5 o mais difícil. É possível verificar, então, que o jornal FSP ficou mais complexo em relação aos índices de leitura e segundo mais complexo em relação ao léxico. Do ponto de vista sintático, ficou mais acessível que o DG, jornal popular.

O jornal ZH é o segundo mais difícil em relação à leitura, terceiro mais difícil em relação ao léxico e empata com o DG quanto à sintaxe.

O DG é o texto mais simples do ponto de vista lexical, empata com o JM na leitura e com o jornal ZH na sintaxe.

O JM ficou como o texto mais simples do ponto de vista sintático, segundo mais simples do ponto de vista do léxico e empata com o DG em termos de leitura.

E o CE ficou como mais fácil em relação aos índices de leitura, e as possíveis razões para tanto já apontamos. Mas o que as colunas amarelas (léxico) e verdes (sintaxe) deixam patente no gráfico é o fato de este ser o *corpus* mais complexo em relação ao léxico e à sintaxe.

Tal resultado confirma a Hipótese 1 de pesquisa colocada no Capítulo 2 deste trabalho, ou seja: **o texto da maioria das Sentenças dos JECs do PJRS apresenta uma configuração lexical, morfosintática e terminológica pouco coincidente com padrões do Português Popular Escrito (PPE). E essa configuração sintática e lexical pode dificultar sua inteligibilidade para pessoas de escolaridade limitada e pouca experiência em leitura.**

Diante dessa configuração do CE, cremos que processos de simplificação lexical e sintática desses textos, como fizemos no Estudo-Piloto 2, podem aumentar a acessibilidade textual. Com isso, podem aumentar a compreensão do que neles está escrito por parte do leitor a que se destinam: o cidadão que busca nos JECs a solução para uma necessidade. A complexidade do léxico desponta como um primeiro fator a exigir alguma intervenção, o que também se aplica, em seguida, à sintaxe. Acreditamos que assim, com uma descrição multifatorial, cientificamente embasada, sobre a tessitura desse tipo de texto, se poderá dar um passo a mais para realmente atender aos requisitos da simplicidade e informalidade expressos no art. 2º da Lei n. 9.099, de 1995. O diagnóstico linguístico, assim, busca ultrapassar as impressões mais subjetivas e pretende colocar pontos focais para uma intervenção transformadora.

Até aqui comparamos o CE com os *corpora* de contraste com ênfase nos aspectos sintáticos. Ainda que aspectos lexicais também tenham sido contemplados nas análises aqui expostas, no próximo capítulo, ampliaremos o espectro no estudo do léxico de nossos *corpora*. Vamos apresentar como se comporta o léxico do CE em relação aos *corpora* de contraste e de referência relacionados no capítulo sobre os procedimentos metodológicos.

9 PADRÕES LEXICAIS DOS *CORPORA* ANALISADOS

Neste capítulo, atendendo aos objetivos estabelecidos no Capítulo 3 deste trabalho, seguimos em mais uma etapa de descrição e comparação entre o CE e os *corpora* de contraste e de referência. Desta vez, nosso foco são os padrões lexicais.

A comparação lexical que estabelecemos aqui tem a intenção de verificar o quanto os padrões lexicais do CE são próximos ou distantes dos padrões lexicais dos textos dos *corpora* CCA, CCB, CCP, CR1 e CR2. Havendo discrepâncias que evidenciem desvios significativos, isso poderá ser considerado um fator que contribui para a dificuldade de compreensão dos textos do CE.

Após, com base nessas constatações e nas apresentadas nos Capítulos 6, 7 e 8, proporemos alternativas de reescrita simplificada para algumas das situações consideradas complexas ao leitor que este trabalho tem por foco: um cidadão brasileiro com pouca escolaridade e pouca experiência em leitura.

Retomamos aqui que esta pesquisa se vale dos pressupostos da Linguística de *Corpus*. Com base nesses pressupostos, é somente por meio de comparações e contrastes que se pode delimitar o que é verdadeiramente peculiar em um *corpus* de linguagem específico, no caso o nosso CE.

A seguir, na Tabela 9.1, repetimos a Tabela 5.6, com os dados de *tokens* e *types* dos *corpora* aqui analisados.

Tabela 9.1 – Número de *Tokens* e *Types* do CE, dos *Corpora* de Contraste e de Referência

<i>CORPORA</i>		Quantidade de <i>tokens</i>	<i>Types</i>
CE		542.836	14.665
CCA	FSP	544.246	30.333
	ZH	507.075	28.630
CCB	DG	496.712	30.196
	JM	491.667	27.990
CCP		684.799	32.138
CR1		2.460.268	82.136
CR2		3.896.392.719 ¹⁰⁵	

¹⁰⁵ Registramos apenas o número de palavras, conforme observação da nota de rodapé 36.

Os dados relativos aos *types* e *tokens* da Tabela 9.1 foram extraídos após a submissão de cada um dos *corpora* à ferramenta *AntConc3.2.1w*. As listas geradas por essa ferramenta apresentam todas as formas das palavras. É o que ocorre, por exemplo, com a palavra **autor** no CE. A forma **autor** ocorre 1891 vezes; **autores**, 317; **autora**, 2674; e **autoras**, 19 vezes.

As análises aqui apresentadas são relativas aos *types*, uma vez que eles representam as palavras diferentes que ocorrem no texto. E o nosso foco aqui foi justamente verificar as peculiaridades lexicais do CE em relação aos demais *corpora*.

Para melhor compreensão do procedimento adotado nas comparações que estabelecemos neste capítulo, apresentamos como exemplo a comparação entre os 30 primeiros *types* da lista de palavras do CE com os 30 primeiros *types* da lista do jornal Zero Hora.

Tabela 9.2 – Exemplo de Comparação entre Listas de Palavras

	30 PRIMEIROS TYPES DO CE	FREQUÊNCIA	30 PRIMEIROS TYPES DO JORNAL ZH	FREQUÊNCIA
1.	a	22165	de	25846
2.	do	15843	a	18953
3.	que	13841	o	17245
4.	da	13668	que	12278
5.	o	13657	e	11746
6.	e	10273	do	10228
7.	em	8591	da	8869
8.	se	6231	em	8019
9.	não	5835	para	7309
10.	no	5688	no	5516
11.	por	4341	com	5382
12.	ao	4208	na	4575
13.	para	3951	é	4321
14.	na	3421	os	4117
15.	com	3406	não	4116
16.	de	3297	um	3717
17.	os	3270	uma	3093
18.	dos	3065	as	3054
19.	autora	2764	por	3030
20.	à	2736	se	2980
21.	pela	2735	mais	2408
22.	lei	2455	ao	2328

23.	é	2449	dos	2258
24.	as	2385	como	2071
25.	art.	2376	foi	1922
26.	nos	2275	das	1753
27.	danos	2159	mas	1703
28.	como	2133	à	1573
29.	parte	2069	pelo	1451
30.	pelo	2069	presidente	1368

A primeira coluna da esquerda para a direita na Tabela 9.2, está numerada segundo a frequência de cada *type* no texto. As palavras que aparecem na primeira linha, com o número 1 nesta coluna, são as que ocorreram mais vezes nos textos. Assim, o *type a* foi o que mais ocorreu no CE, e o *type de* foi o que mais ocorreu no jornal Zero Hora. A coluna denominada de Frequência indica o número de ocorrências de cada palavra no texto. O **a**, primeiro *type* da tabela do CE, apareceu 22.165 vezes no CE, e o **de** ocorreu 25.846 vezes no texto do jornal Zero Hora.

Nas duas listas exemplificadas na Tabela 9.2, observamos um fenômeno comum a qualquer lista de palavras de um texto: as palavras com maior frequência, e por isso as primeiras a aparecerem nas listas, são as gramaticais. Biderman (1998), com base em um *corpus* de 5 milhões de palavras do português do Brasil escrito (1950-1995) – composto por literatura romanesca, dramática, técnico-científica, jornalística e oratória –, verificou que as palavras gramaticais são as mais frequentes. Ela também verificou que 80% de qualquer texto do português é constituído por pouco mais de 1000 palavras mais frequentes (artigos, preposições, pronomes, conjunções, advérbios e numerais) e algumas palavras que ela chama de plenas (substantivos, adjetivos e verbos).

No caso da lista do CE, a primeira palavra lexical é **autora**, na 19ª colocação. Na lista do jornal ZH, a primeira palavra lexical ocorre na 13ª colocação: o verbo ser - **é**. Antes delas, ocorrem apenas palavras funcionais ou gramaticais.

Além, disso, se somarmos o total de ocorrências dessas primeiras 20 palavras em cada *corpus*, temos os seguintes dados:

- no CE, a soma gera 173.356 ocorrências desses 20 *types*, que equivalem a 32% do total de *tokens*, 542.836.
- no jornal ZH, a soma gera 183.229 ocorrências desses 20 *types*, o que equivale a 36% do total de *tokens*, 507.075.

- Ainda em relação à Tabela 9.2, algumas células estão preenchidas com a cor verde, e outras não. Nas células preenchidas com a cor verde estão as palavras que ocorrem nas duas listas. Por exemplo, o **a**, primeiro *type* no CE, ocorre também no jornal ZH, porém na segunda colocação. O **de**, primeiro colocado no jornal ZH, ocorre também no CE, porém na 16ª colocação.

Para identificarmos as ocorrências repetidas das listas de palavras aqui apresentadas, valemo-nos do programa *Microsoft Office Excel*, que, entre tantos recursos, oferece o de destacar elementos duplicados em uma relação de itens que nele são colocados. Após a extração das listas geradas pelo AntConc3.2.1w, salvamos cada uma delas no *Microsoft Office Excel* e, para a comparação, colamos as listas em contraste uma abaixo da outra, cada uma com uma cor de fonte diferente. No caso do CE e do jornal ZH, por exemplo, colocamos as palavras do CE na cor Automático (preta), e as do jornal ZH em vermelho. Após, ativamos, pelo comando do *Microsoft Office Excel*, a identificação dos elementos duplicados com o preenchimento das células respectivas na cor verde. O resultado está exemplificado na Tabela 9.2.

Após essa etapa, geramos três listas: uma só com os elementos duplicados nas duas listas, outra só com os elementos exclusivos (não duplicados) do CE, e outra só com os elementos exclusivos (não duplicados) no *corpus* em comparação. Exemplificamos, nos Quadros 9.1 e 9.2 com as listas do CE e do jornal ZH trazidas na Tabela 9.2.

Quadro 9.1 – Lista de Palavras Duplicadas entre os 30 primeiros *types* do CE e do Jornal ZH

a	de	no
a	do	no
à	do	o
à	dos	o
ao	dos	os
ao	e	os
as	e	para
as	é	para
com	é	pelo
com	em	pelo
como	em	por
como	na	por
da	na	que
da	não	que
de	não	se
		Se

O Quadro 9.1 apresenta 46 palavras duplicadas. Colocamos em ordem alfabética para melhor visualização. As palavras em vermelho correspondem à lista do jornal ZH, e as palavras em preto correspondem ao CE.

Após essa identificação dos duplicados, geramos as listas de palavras exclusivas de cada *corpus*. É o que consta no Quadro 9.2.

Quadro 9.2 – Lista de Palavras Exclusivas do CE e do Jornal ZH entre os 30 primeiros *types*

Palavras exclusivas do CE	Palavras exclusivas do jornal ZH
<u>autora</u>	um
pela	uma
<u>lei</u>	<u>mais</u>
<u>art.</u>	<u>foi</u>
nos	das
<u>danos</u>	mas
<u>parte</u>	<u>Presidente</u>

As duas listas do Quadro 9.2 geraram um total de 14 palavras exclusivas nos *corpora* em contraste. Aqui aparecem as palavras lexicais. No CE, das 7 palavras, 5 são lexicais (sublinhadas na lista). No jornal ZH, apenas 3 são lexicais (também sublinhadas na lista). Considerando que as palavras lexicais são as que carregam o conteúdo, a lista do CE é mais densa semanticamente que o jornal ZH. Trata-se de um pequeno excerto da lista e, como observamos no Capítulo 8, o fato de haver mais palavras lexicais não pode ser tomado isoladamente. É preciso verificar o quanto essas palavras lexicais são potencialmente conhecidas ou não pelo leitor. De todo modo, é um dado que o exemplo em análise permite observar.

A lista de palavras exclusivas dessa amostra dos dois *corpora* pode ser representada pela seguinte figura.

Figura 9.1 – Exemplo de *Types* do CE e do Jornal ZH



Na Figura 9.1, temos dois círculos: um referente ao CE (na cor mostarda), outro referente ao jornal ZH (na cor azul). Esses círculos se interseccionam. E é nessa intersecção (na cor acinzentada) que constam as palavras duplicadas, exemplificadas com a fonte branca. As palavras exclusivas a cada um dos *corpora* se encontram nos círculos relativos a cada um deles.

Com esses dados, podemos dizer que, do total de 14 palavras exclusivas dessa amostra dos 30 primeiros *types* dos *corpora* em contraste, 50% são exclusivas do CE, e não constam no jornal ZH. Disso podemos inferir que o leitor do jornal ZH terá, em tese, 50% de chance de não compreender o vocabulário empregado nesse pequeno excerto das sentenças aqui analisadas.

Feito o esclarecimento sobre como procedemos nas comparações entre os *corpora* aqui apresentadas, iniciamos a exposição dos resultados relativos às comparações do CE com o *Corpus* de Contraste – CCA.

9.1 CE X CCA

O CCA é composto de dois *corpora*: dos jornais Folha de São Paulo (FSP) e Zero Hora (ZH). Iniciaremos com as comparações do CE com o jornal ZH.

9.1.1 CE X ZH

Adotando o procedimento exemplificado acima à totalidade dos *corpora* do CE e do jornal Zero Hora, obtivemos os seguintes dados, apresentados na Tabela 9.3.

Tabela 9.3 – Dados dos *Types* do CE e do Jornal ZH

<i>Corpus</i>	Total de <i>types</i>	Quantidade de <i>types</i> duplicados		Quantidade de <i>types</i> exclusivos	
		Número bruto	Percentual	Número bruto	Percentual
ZH	28.630	8.888	21%	19.742	46%
CE	14.665	8.943	21%	5774	13%
TOTAL	43.295	17.831	41%	25.516	59%

Os percentuais colocados na Tabela 9.3 referem-se à relação entre o número bruto de *types* duplicados e exclusivos e o total de *types* de ambos os *corpora* (43.295). Os dados da Tabela 9.3, acima, podem ser melhor visualizados na Figura 9.2, a seguir.

Figura 9.2 – Percentuais de *Types* Exclusivos e Duplicados do CE e do Jornal ZH



Para exemplificarmos alguns casos dessas ocorrências, criamos a Tabela 9.4, a seguir, que traz as 20 palavras lexicais mais frequentes de cada conjunto.

Tabela 9.4 – Distribuição dos 20 *Types* mais frequentes na Comparação entre o CE e o Jornal ZH

EXCLUSIVOS CE (13%)		DUPLICADOS (41%)		EXCLUSIVOS ZH (46%)	
FREQ.	TYPE	FREQ.	TYPE	FREQ.	TYPE
791	demandada	9951	não	958	Bolsonaro
750	recursais	6770	é	933	coronavírus
564	opino	3341	ser	631	pandemia
538	recursal	3227	foi	461	crise

514	demandante	2897	mais	429	covid
425	improcedência	2766	autora	308	Jair
422	homologatória	2616	lei	264	governador
413	homologo	2505	parte	261	bilhões
397	intimadas	2180	danos	259	Moro
372	<i>quantum</i>	2142	caso	214	secretário
370	inominado	2048	ré	209	diretor
325	inversão	1990	valor	173	vamos
305	reparação	1926	sentença	171	distanciamento
297	Fulano	1910	autor	167	combate
291	indenizatório	1689	presidente	166	Guedes
273	apelação	1626	dano	162	chefe
271	advocaticios	1595	pedido	162	mortes
247	súmula	1521	direito	161	agricultura
232	consumidora	1512	recurso	161	leitos
229	<i>re*</i>	1488	morais	159	Entrevista

Na coluna em laranja, constam as primeiras 20 palavras exclusivas mais frequentes do CE. A primeira palavra que aparece é **demandada**, que se refere a uma das partes de uma demanda judicial (de um processo). Em seguida vem **recursais**, adjetivo que acompanha a palavra **turmas**, na expressão *Turmas Recursais*. Esse nome refere-se aos grupos que julgam os recursos dos JECs, conforme mencionamos na seção 5.2.1. Em terceiro, temos a forma verbal **opino**, na 1ª pessoa do singular. É a forma como normalmente o juiz leigo conclui o seu parecer (Sentença) sobre o caso em julgamento.

Exemplo:

(9.1) *Ante o exposto, **opino** pela improcedência dos pedidos, com fundamento [...].*

E as demais palavras da lista, em tese, podem ser consideradas termos, porque, ainda que existam fora do mundo jurídico, nas sentenças recebem significado especializado. É o caso de **inversão** (no jornal ZH não consta), que, nas 325 ocorrências, em 323 está inserida no sintagma terminológico **inversão do ônus da prova**. Esta expressão significa que a responsabilidade de provar algo no processo não é daquele que alega, mas daquele contra quem se alega algo, ou seja, de quem está sendo processado.

Exemplo:

(9.2) *Em razão da **inversão do ônus da prova**, ao réu cabe a comprovação da existência de [...].*

Chama também a atenção a presença de *quantum*, em latim, que significa **quanto**. Nas Sentenças aqui analisadas é empregado com o significado de **quantia**. Das 372 ocorrências, em 204 vem acompanhado do adjetivo **indenizatório**, formando um sintagma: **quantum indenizatório**. É uma construção híbrida, pois relaciona palavras de duas línguas numa expressão só. Tal construção se repete quanto aparece acompanhado dos adjetivos **mantido**, **fixado**: **quantum mantido**, **quantum fixado**. Houve 5 ocorrências da expressão totalmente em latim **quantum debeat**, que significa *quantia devida*.

A última palavra da lista – *re* – ocorre 226 vezes no interior da expressão **in re ipsa**, que significa *da própria coisa, do próprio fato*, algo que é presumido, que não precisa de prova. Um dano **in re ipsa** é um dano que não precisa ser provado, porque é presumido. Um exemplo é a inscrição indevida do nome de alguém em cadastros de inadimplentes. Se não há razão para essa inscrição, o dano causado à pessoa que foi inscrita é admitido como real.

Nas colunas em verde, estão as palavras duplicadas. A primeira ocorrência é o advérbio **não**¹⁰⁶. As próximas três primeiras ocorrências são formas do verbo **ser**: *é*, *ser* e *foi*. Este verbo é um dos verbos mais usados na Língua Portuguesa, e aqui os dados parecem confirmar que realmente assim é. Em sexto lugar, encontramos a palavra **autora**, que, no Quadro 9.2, foi o *type* mais frequente do CE. Ocorre que, nesse caso, a comparação foi com as primeiras 30 palavras mais frequentes, e aqui se faz com a totalidade dos *corpora*. Mesmo ocorrendo apenas 2 vezes no jornal ZH, a palavra **autora** passa a integrar a lista dos duplicados.

Muitas das palavras da lista de duplicados, quando empregadas no contexto das sentenças, assumem uma especificidade de sentido que podem lhe atribuir o caráter de termo. Este é um fenômeno muito comum nas linguagens especializadas, principalmente no Direito. É o caso, por exemplo, de **autora e parte** no Quadro 9.3, a seguir.

Quadro 9.3 – Exemplos de Palavras Duplicadas no CE e no Jornal ZH

CE	ZERO HORA
Aduziu a Autora que somente adquiriu os móveis e utensílios da Requerida porque pretendia dar continuidade ao negócio.	Também é autora do livro de crônicas Tudo Tem uma Primeira Vez.

¹⁰⁶ Comentaremos sobre a ocorrência do advérbio “não” após a exposição dos jornais JM e DG.

[...] o valor da indenização deve ter o condão de punir e educar a parte ofensora (viés punitivo-pedagógico da responsabilidade civil)	O secretário passou parte do dia de ontem na sede da Federação da Agricultura do Estado (FARSUL).
---	--

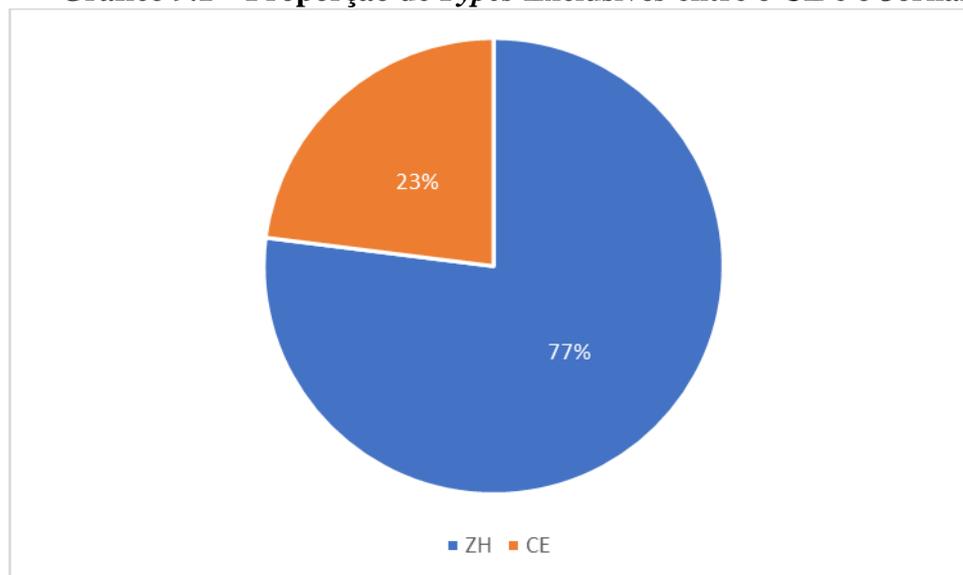
Na coluna em azul, constam as palavras exclusivas do jornal ZH. Como se vê, elas refletem o período das publicações extraídas do jornal. Em 2019, o Brasil estava no segundo ano do mandato de **Bolsonaro** (palavra lexical mais frequente). E, em 2020, iniciou a **pandemia** (terceira palavra lexical mais frequente) do **coronavírus** (segunda palavra lexical mais frequente).

Após essa visão geral do conjunto de palavras do CE e do ZH, passamos a analisar a relação entre os *types* exclusivos que neles se encontram. A Tabela 9.5 apresenta os dados relativos a essa comparação.

Tabela 9.5 – *Types* Exclusivos do CE e do Jornal ZH

CORPUS	Quantidade de <i>types</i> exclusivos	Percentual de <i>types</i> exclusivos
ZH	19.742	77%
CE	5774	23%
TOTAL	25.516	100%

Os percentuais colocados na última coluna da Tabela 9.5 são referentes à quantidade de *types* exclusivos de cada *corpus* em relação ao total de *types* exclusivos, conforme a última linha da segunda coluna (25.516). O Gráfico 9.1 ilustra esse cálculo.

Gráfico 9.1 – Proporção de *Types* Exclusivos entre o CE e o Jornal ZH

Pelo Gráfico 9.1, podemos entender que, do total de *types* exclusivos do CE e do Jornal Zero Hora juntos, 23% referem-se apenas a palavras que se encontram nas Sentenças, e 77% a palavras que se encontram apenas no jornal ZH. Considerando que nosso foco está nas Sentenças, **podemos pressupor que o leitor do jornal ZH terá, em tese, problemas em compreender 23% do vocabulário que se encontra nas Sentenças aqui analisadas.**

Entretanto, se relacionarmos o número de palavras exclusivas do CE (5.774), conforme Tabela 9.5, com o número total de *types* que este mesmo *corpus* apresenta (14.665), conforme Tabela 9.3, chegaremos a um percentual de 39,23%. Assim, com base nessa relação, é **também possível pressupor que o leitor do jornal ZH terá, em princípio, chance de não compreender aproximadamente 40% do vocabulário das sentenças aqui analisadas.**

Cabe referir que, como já observamos quando apresentamos este *corpus*, na seção 5.1.2, que, segundo a Wikipedia, o jornal Zero Hora é um dos maiores jornais de circulação no país.

Sabendo-se que os jornais de referência são, em tese, destinados às classes mais favorecidas da sociedade e, em princípio, com maior escolaridade e experiência em leitura, o fato de 40% do vocabulário das sentenças aqui analisadas não serem do conhecimento desse público é preocupante.

Como o CCA é também composto por outro jornal de referência, o jornal Folha de São Paulo (FSP), veremos primeiro como se dão as comparações com este outro *corpus*. Feito isso, teceremos outros comentários ao colocado no parágrafo anterior.

9.1.2 CE X FSP

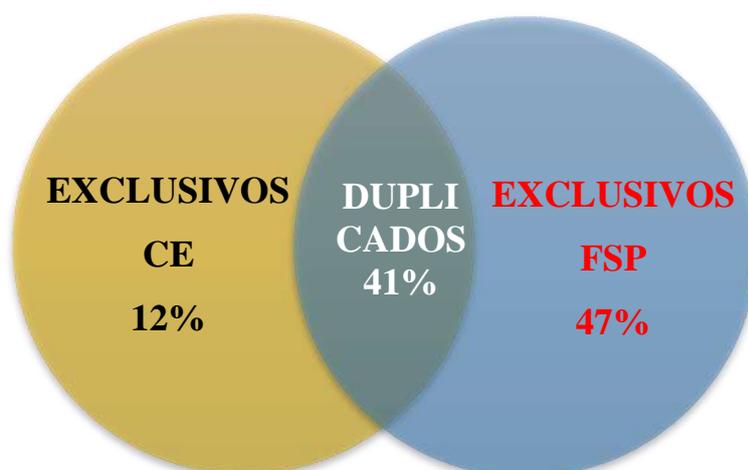
Após compararmos as listas de palavras do CE com o jornal FSP, chegamos aos dados apresentados na Tabela 9.6.

Tabela 9.6 – Dados dos *Types* do CE e do Jornal FSP

<i>Corpus</i>	Total de <i>types</i>	Quantidade de <i>types</i> duplicados		Quantidade de <i>types</i> exclusivos	
		Número bruto	Percentual	Número bruto	Percentual
FSP	30.333	9.117	20%	21.216	47%
CE	14.665	9.120	20%	5545	12%
TOTAL	44.998	18.237	41%	26.761	59%

Os percentuais colocados na Tabela 9.6 referem-se à relação entre o número bruto de *types* duplicados e exclusivos e o total de *types* de ambos os *corpora*. Os dados da Tabela 9.6, acima, podem ser melhor visualizados na Figura 9.3, a seguir.

Figura 9.3 – Percentuais de *Types* Exclusivos e Duplicados do CE e do Jornal FSP



Para exemplificarmos alguns casos dessas ocorrências, criamos a Tabela 9.7, a seguir, que traz as 20 palavras lexicais mais frequentes de cada conjunto.

Tabela 9.7 – Distribuição dos 20 *types* mais frequentes na comparação entre o CE e o Jornal FSP

EXCLUSIVOS CE (13%)		DUPLICADOS (41%)		EXCLUSIVOS FSP (46%)	
FREQ.	TYPE	FREQ.	TYPE	FREQ.	TYPE
874	honorários	10676	não	2281	Bolsonaro
818	custas	3504	ser	1028	coronavírus
791	demandada	3498	foi	553	pandemia
750	recursais	2770	autora	492	crise
564	opino	2704	mais	480	Moro
514	demandante	2510	parte	464	covid
450	CPC (Código de Processo Civil)*	2500	presidente	445	Jair
433	CDC (Código de Defesa do Consumidor)*	2394	art.	298	mortes
425	improcedência	2331	caso	285	deputado
422	homologatória	2189	danos	264	Mandetta
413	homologo	2052	ré	261	PSL
397	intimadas	2016	valor	254	Lula
393	leigo	1965	são	244	EUA
388	produza	1956	autor	238	bilhões
372	<i>quantum</i>	1955	sentença	210	ministros
370	inominado	1627	dano	210	PF
316	IGP (IGP-M – Índice Geral de Preços do Mercado)*	1521	recurso	190	quarentena
291	indenizatório	1495	morais	187	governadores
232	consumidora	1452	autos	186	governador
229	<i>re</i> *	1445	governo	186	leitos

Na coluna em laranja, constam as primeiras 20 palavras exclusivas mais frequentes do CE. A primeira palavra – **honorários** – se refere à remuneração que um

profissional liberal recebe por um serviço prestado. Como muitos temas das Sentenças são relativos a serviços prestados e como é frequente a expressão **sem custas e honorários** ao final da sentença, é plausível a sua maior ocorrência no CE. Tal como ocorreu com a comparação com o jornal ZH, as demais palavras exclusivas do CE são específicas do mundo jurídico: **custas, demandada, recursais, demandante**, etc.

Quatro palavras da lista exclusiva do CE estão com asterisco. Três são siglas e, apesar de abreviarem vários *types*, são consideradas 1 *type* pela ferramenta AntConc. E assim consideramos também. Colocamos entre parênteses o que significam. No caso do IGP-M, a grafia IGPM (70 ocorrências) não foi computada no total de 316, conforme a Tabela 9.7.

A sigla CDC, que significa Código de Defesa do Consumidor, nas 433 ocorrências, em apenas 7 veio acompanhada do nome a que se refere por extenso. As outras duas – CPC e IGP-M –, em todas as ocorrências, vieram desacompanhadas do nome que representam por extenso.

O CDC é uma lei muito utilizada para fundamentar as decisões dos Juizados Especiais Cíveis, e o fato de ela ocorrer quase sempre desacompanhada do nome por extenso permite pensar que haja o pressuposto de que ela seja de conhecimento de todos os seus leitores. E parece não ser bem o caso, pois ela consta exclusivamente no CE.

No Brasil, de acordo com o Princípio da Obrigatoriedade das Leis expresso no art. 3º da Lei de Introdução às Normas do Direito Brasileiro (LINDB), ninguém pode ser desculpado por não cumprir a lei alegando que não a conhece. Mas aqui é diferente. Estamos tratando apenas do emprego da sigla referente a uma lei.

Quanto à última palavra – *re* -, já comentamos sobre ela na Tabela 9.4.

Nas colunas em verde, estão as palavras duplicadas. Em comparação com os duplicados da relação do CE com o jornal ZH, há apenas 6 palavras diferentes: **art., direito, é, governo, lei e são**. O primeiro *type* mais frequente continua sendo o **não**¹⁰⁷. Os dois próximos *types* mais frequentes continuam sendo formas do verbo **ser**, e **autora** passou para a quarta colocação.

Na coluna em azul, constam as palavras exclusivas do jornal FSP. Tal como ocorreu com o jornal ZH, as palavras exclusivas do jornal FSP também refletem o período das publicações extraídas do jornal. Em 2019, o Brasil estava no segundo ano do mandato de Bolsonaro. Como ocorreu com o jornal ZH, **Bolsonaro** foi a palavra lexical mais

¹⁰⁷ Ver nota n. 2

frequente. E, em 2020, iniciou a **pandemia** (segunda palavra lexical mais frequente) do **coronavírus** (terceira palavra lexical mais frequente).

Quanto a essa semelhança nas palavras empregadas pelos jornais ZH e FSP, cumpre referir que, segundo Amaral (2006, p. 55), o “jornalismo de referência fala, sobretudo, com o leitor interessado no mundo público”. Seria, então, uma das razões para aparecerem palavras ligadas à esfera política e sanitária do momento das notícias.

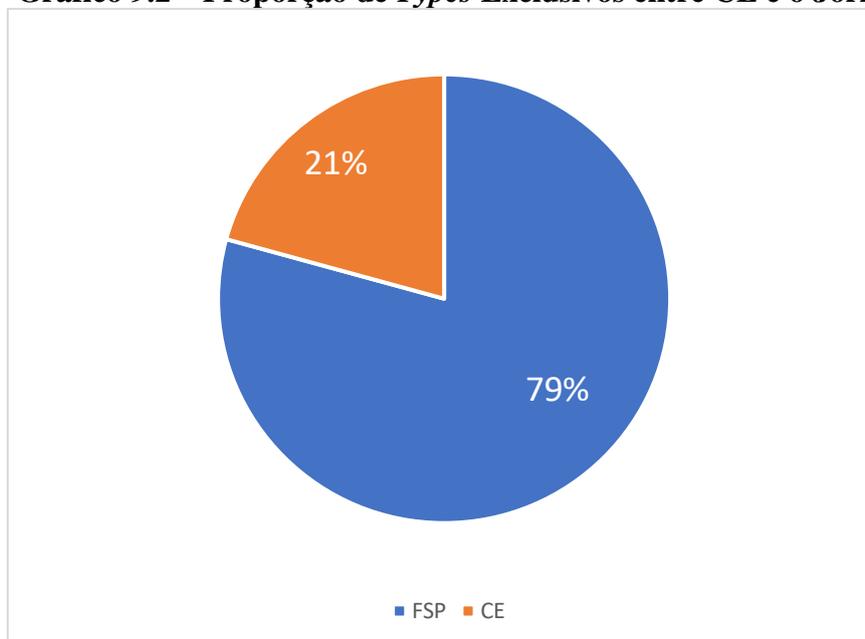
Passamos, agora, a analisar a relação entre os *types* exclusivos que se encontram no CE e no jornal FSP. A Tabela 9.8 apresenta os dados relativos a essa comparação.

Tabela 9.8 – *Types* Exclusivos do CE e do Jornal FSP

<i>CORPUS</i>	Quantidade de <i>types</i> exclusivos	Percentual de <i>types</i> exclusivos
FSP	21.216	79%
CE	5545	21%
TOTAL	26.761	100%

Os percentuais colocados na última coluna da Tabela 9.8 são referentes à quantidade de *types* exclusivos de cada *corpus* em relação ao total de *types* exclusivos, conforme a última linha da segunda coluna (26.761). O Gráfico 9.2 ilustra esse cálculo.

Gráfico 9.2 – Proporção de *Types* Exclusivos entre CE e o Jornal FSP



Pelo Gráfico 9.2, podemos entender que, do total de *types* exclusivos do CE e do Jornal Folha de São Paulo juntos, 21% referem-se apenas a palavras que se encontram

nas Sentenças, e 79% a palavras que se encontram apenas no jornal ZH. Considerando que nosso foco está nas Sentenças, **podemos pressupor que o leitor do jornal FSP terá, em tese, problemas em compreender 21% do vocabulário que se encontra nas Sentenças aqui analisadas.**

Entretanto, se relacionarmos o número de palavras exclusivas do CE (5.545), conforme Tabela 9.8, com o número total de *types* que este mesmo *corpus* apresenta (14.665), conforme Tabela 9.6, chegaremos a um percentual de 37,81%. Assim, com base nessa relação, **é também possível pressupor que o leitor do jornal FSP terá, em princípio, chance de não compreender aproximadamente 40% do vocabulário das sentenças aqui analisadas.**

Os percentuais mencionados nos dois últimos parágrafos assemelham-se com os do jornal ZH. Ou seja, os leitores desses jornais têm a probabilidade de não compreender no mínimo 21% do léxico que se encontra nas sentenças. E, se analisarmos apenas o léxico das sentenças, esse percentual aumenta para em torno de 40%.

Os jornais de referência, como já mencionamos no Capítulo 5, são dirigidos às classes A e B, que têm um nível de escolaridade mais alto. Segundo Amaral (2006, p. 30), entre os leitores do jornal Folha de São Paulo, de acordo os dados do Perfil do leitor 2000/Datafolha), 47% tinham nível superior completo, 13% curso de pós-graduação, 53% com renda individual de até 15 salários mínimos e 35% com renda superior a 30 salários mínimos.

De acordo com dados do IBGE (2020), as classes econômicas com rendimento superior (Classe A - acima de 20 salários mínimos; classe B – de 10 a 20 salários mínimos) é que apresentam percentuais mais altos entre os brasileiros com 15 anos ou mais que concluíram o Ensino Médio (33,7%) e o Ensino Superior (36,9%). Nesse sentido, segundo dados relativos ao ano de 2019 do Anuário Brasileiro de Educação Básica¹⁰⁸, o percentual de alunos que concluíram a 3ª série do Ensino Médio com aprendizagem adequada em Língua Portuguesa foi de 57,6% nas classes econômicas mais altas, e 28,2% nas mais baixas.

Além disso, segundo os dados do INAF (2018) já mencionados neste trabalho, das pessoas que completaram o Ensino Superior ou o fizeram de forma incompleta, apenas 34% encontram-se no nível pleno de alfabetismo.

¹⁰⁸ Anuário Brasileiro de Educação Básica. Disponível em: <https://www.moderna.com.br/anuario-educacao-basica/2021/brasil-principais-dados-e-indicadores-de-desigualdade.html>

Diante desse panorama, o fato de 40% do vocabulário das sentenças ser desconhecido desse perfil de leitor deixa patente o quanto esses textos dos JECs estão distanciados do seu principal destinatário: o cidadão que recorre a esses Juizados para solucionar um problema que lhe aflige.

Vejamos, agora, como se dá a relação entre os jornais populares do CCB – Diário Gaúcho e Jornal Massa.

9.2 CE X CCB

O CCB é composto de dois *corpora*: dos jornais Massa (JM) e Diário Gaúcho (DG). Iniciaremos com as comparações do CE com o JM.

9.2.1 CE X JM

Após compararmos as listas de palavras do CE com o JM, chegamos aos dados apresentados na Tabela 9.9.

Tabela 9.9 – Dados dos *Types* do CE e do JM

<i>Corpus</i>	Total de <i>types</i>	Quantidade de <i>types</i> duplicados		Quantidade de <i>types</i> exclusivos	
		Número bruto	Percentual	Número bruto	Percentual
JM	27.990	8.107	19%	19.883	47%
CE	14.665	8.106	19%	6.559	15%
TOTAL	42.655	16.213	38%	26.442	62%

Os percentuais colocados na Tabela 9.9 referem-se à relação entre o número bruto de *types* (duplicados e exclusivos) e o total de *types* de ambos os *corpora*. Os dados da Tabela 9.9, acima, podem ser melhor visualizados na Figura 9.4, a seguir.

Figura 9.4 - Percentuais de *Types* Exclusivos e Duplicados do CE e do JM



Para exemplificarmos alguns casos dessas ocorrências, criamos a Tabela 9.10, a seguir, que traz as 20 palavras lexicais mais frequentes de cada conjunto.

Tabela 9.10 – Distribuição dos 20 *Types* mais Frequentes na Comparação entre o CE e o JM

EXCLUSIVO CE (15%)		DUPLICADOS (38%)		EXCLUSIVO FSP (47%)	
FREQ.	TYPE	FREQ.	TYPE	FREQ.	TYPE
791	demandada	9069	não	1275	Bahia
750	recursais	4792	texto	1019	Salvador
564	opino	3595	foi	734	pandemia
538	recursal	3070	ser	712	covid
514	demandante	2768	autora	696	coronavírus
450	CPC (Código de Processo Civil)*	2577	lei	427	time
425	improcedência	2393	parte	354	jogos
422	homologatória	2300	mais	293	campeonato
413	homologo	2174	danos	246	futebol
397	intimadas	2048	ré	244	baiana
393	leigo	2000	caso	244	tricolor
372	<i>quantum</i>	1917	autor	232	vacina
370	inominado	1916	sentença	224	leão
352	inexistência	1893	valor	218	baiano
327	juizado	1722	nº	218	massa

316	IGP-M (Índice Geral de Preços do Mercado)*	1623	dano	216	competição
310	faturas	1540	direito	213	copa
297	Fulano	1525	pedido	208	leitos
291	indenizatório	1504	morais	189	ACM (Antônio Carlos Magalhães Neto)
273	apelação	1493	recurso	189	carnaval

Na coluna em laranja, constam as primeiras 20 palavras exclusivas mais frequentes do CE. A primeira palavra que aparece é **demandada**. Como já vimos nos jornais ZH e FSP, esta palavra continua sendo uma das primeiras no *ranking* exclusivo do CE. Além disso, as 5 primeiras palavras do JM são as 5 primeiras palavras lexicais mais frequentes no jornal ZH. E, igualmente, tal como ocorreu com a comparação com o grupo CCA, as demais palavras exclusivas do CE são específicas do mundo jurídico: **CPC**, **inominado**, *quantum*, etc.

Aparecem novamente 2 siglas, com asteriscos. Como já mencionamos, apesar de abreviarem vários *types*, são consideradas 1 *type* pela ferramenta AntConc. E assim consideramos também. Colocamos entre parênteses o que significam.

Nas colunas em verde, estão as palavras duplicadas. A primeira palavra mais frequente continua sendo o advérbio **não**¹⁰⁹. E a segunda mais frequente, diferentemente dos jornais anteriores, foi **texto**. Depois dela ocorrem **foi**, **ser** e **autora**, que também se mantiveram entre as seis primeiras do *ranking* nos jornais anteriores.

Na coluna em azul, constam as palavras exclusivas do jornal JM. Diferentemente dos jornais ZH e FSP, as primeiras palavras mais frequentes são **Bahia** e **Salvador**, local onde é produzido este jornal. Apesar de suas matérias terem sido também extraídas entre os anos de 2019 e 2020, as primeiras palavras não se relacionam a políticos, como aconteceu com os jornais anteriores. A primeira menção, na 19ª colocação, é **ACM**, que se refere ao Prefeito de Salvador Antônio Carlos Magalhães Neto.

O tema da pandemia, porém, começa a aparecer a partir da terceira palavra mais frequente: **pandemia** (3ª), **covid** (4ª) e **coronavírus** (5ª). Depois são palavras relacionadas a futebol, a jogos, e a vigésima é **carnaval**. Tais frequências de palavras lexicais parecem confirmar o propósito editorial dos jornais populares, como é o caso do

¹⁰⁹ Quando abordarmos esses itens no DG, abordaremos a ocorrência do advérbio **não** nos jornais aqui analisados.

JM. Ou seja, o objetivo é priorizar matérias de entretenimento em vez de matérias de cunho político e econômico. É o que diz Amaral (2006, p. 64-5):

Interessam aos leitores das classes C, D e E temas que digam respeito ao seu cotidiano, especialmente atendimento à saúde, mercado de trabalho, segurança pública, televisão, futebol e as matérias conhecidas como de interesse humano, que contam os dramas cotidianos da população. Também interessa o local em detrimento do nacional.

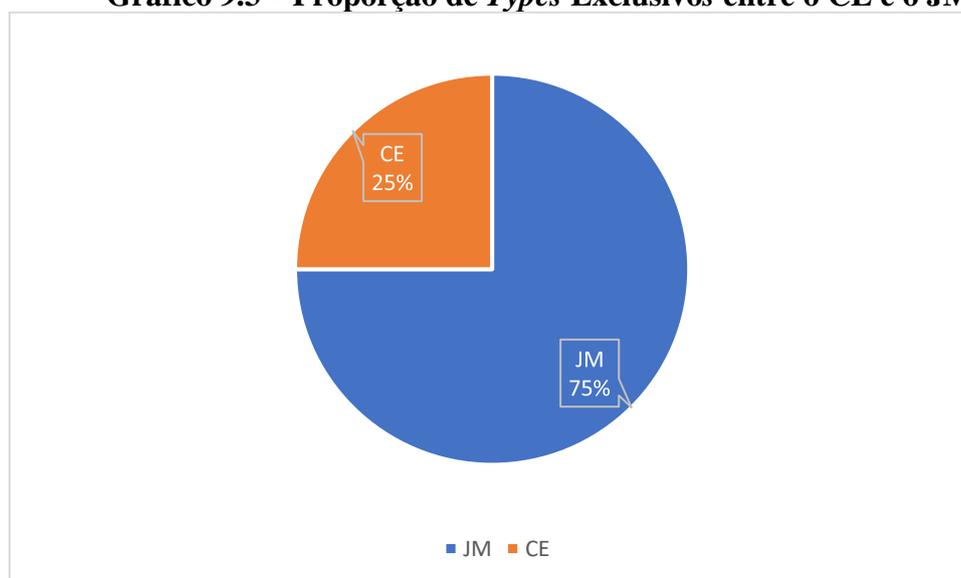
Tendo esse panorama inicial do léxico empregado no CE em comparação com o JM, passamos, agora, a analisar a relação entre os *types* exclusivos que se encontram no CE e no jornal JM. A Tabela 9.11 apresenta os dados relativos a essa comparação.

Tabela 9.11 – *Types* Exclusivos do CE e do JM

CORPUS	Quantidade de <i>types</i> exclusivos	Percentual de <i>types</i> exclusivos
JM	19.883	75%
CE	6.559	25%
TOTAL	26.442	100%

Os percentuais colocados na última coluna da Tabela 9.11 são referentes à quantidade de *types* exclusivos de cada *corpus* em relação ao total de *types* exclusivos, conforme a última linha da segunda coluna (26.442). O Gráfico 9.3 ilustra esse cálculo.

Gráfico 9.3 – Proporção de *Types* Exclusivos entre o CE e o JM



Pelo Gráfico 9.3, podemos entender que, do total de *types* exclusivos do CE e do JM juntos, 25% referem-se apenas a palavras que se encontram nas Sentenças, e 75% a

palavras que se encontram apenas no JM. Considerando que nosso foco está nas Sentenças, podemos pressupor que **o leitor do JM terá, em tese, problemas em compreender 25% do vocabulário que se encontra nas Sentenças aqui analisadas.**

Entretanto, se relacionarmos o número de palavras exclusivas do CE (6.559), conforme Tabela 9.11, com o número total de *types* que este mesmo *corpus* apresenta (14.665), conforme Tabela 9.9, chegaremos a um percentual de 44,72%. Assim, com base nessa relação, **é também possível pressupor que o leitor do JM terá, em princípio, chance de não compreender aproximadamente 45% do vocabulário das sentenças aqui analisadas.**

Passamos aos dados relativos à comparação entre o CE e o DG, segundo integrante do CCB.

9.2.2 CE X DG

Após compararmos as listas de palavras do CE com o DG, chegamos aos dados apresentados na Tabela 9.12.

Tabela 9.12 – Dados dos *Types* do CE e do DG

<i>Corpus</i>	Total de <i>types</i>	Quantidade de <i>types</i> duplicados		Quantidade de <i>types</i> exclusivos	
		Número bruto	Percentual	Número bruto	Percentual
DG	30196	7.941	18%	22255	50%
CE	14.665	7.868	17%	6.797	15%
TOTAL	44861	15.809	35%	29052	65%

Os percentuais colocados na Tabela 9.12 referem-se à relação entre o número bruto de *types* (duplicados e exclusivos) e o total de *types* de ambos os *corpora*. Os dados da Tabela 9.12, acima, podem ser melhor visualizados na Figura 9.5, a seguir.

Figura 9.5 – Percentuais de *Types* Exclusivos e Duplicados do CE e do DG



Para exemplificarmos alguns casos dessas ocorrências, criamos a Tabela 9.13, a seguir, que traz as 20 palavras lexicais mais frequentes de cada conjunto.

Tabela 9.13 – Distribuição dos 20 *Types* mais Frequentes na Comparação entre o CE e o DG

EXCLUSIVO CE (15%)		DUPLICADOS (35%)		EXCLUSIVO DG (50%)	
FREQ.	TYPE	FREQ.	TYPE	FREQ.	TYPE
1620	dano	9399	não	2284	subtítulo
1405	cível	6146	é	656	Grêmio
1092	requerida	3525	foi	583	time
874	honorários	3310	texto	558	amor
791	demandada	3036	ser	467	Inter
750	recursais	2767	autora	336	Renato
538	recursal	2545	lei	288	gols
514	demandante	2379	art.	279	brasileirão
434	CDC (Código de Defesa do Consumidor)*	2335	parte	257	futebol
430	indevida	2314	autor	243	jogador
425	improcedência	2247	anos	239	jogadores
422	homologatória	2245	mais	230	jogos
413	homologo	2167	danos	228	RBS
397	intimadas	2048	ré	222	atletico
393	leigo	1935	caso	218	show
388	produza	1913	sentença	212	ingressos
372	<i>quantum</i>	1885	valor	206	libertadores
370	inominado	1709	nº	206	Roth
327	inciso	1508	morais	195	colorado
310	faturas	1482	recurso	191	homens

Na coluna em laranja, constam as primeiras 20 palavras exclusivas mais frequentes do CE. A primeira palavra que aparece é **dano**. Quando uma pessoa busca os JECs, é porque acha que teve algum dano, algum prejuízo. **Dano**, então, corresponde a uma das temáticas principais da sentença. **Demandada**, primeira palavra exclusiva do CE na relação com o JM, aqui vai para o 5º lugar. Mesmo assim, continua sendo uma das primeiras no *ranking* exclusivo do CE. Quanto às demais palavras, seguem sendo quase todas específicas do mundo jurídico: **recursal**, **recursais**, **demandante**, etc.

Nas colunas em verde, estão as palavras duplicadas. A primeira palavra mais frequente, assim como nos jornais anteriores, foi o advérbio **não**. Como já vimos no Estudo-Piloto 1 e nas análises do Capítulo 8, palavras de negação tendem a deixar o texto mais complexo.

Retomamos aqui o *ranking* dos jornais em relação ao CE quanto à incidência de negação colocado no Capítulo 8: o JM foi o mais fácil, seguido do DG, do jornal ZH, do jornal FSP e do CE. Ou seja, o CE ficou como o mais difícil, e o JM como o mais fácil. E a frequência apenas do advérbio **não** confirma esse *ranking*: é o 10º *type* mais frequente no CE; o 12º no jornal FSP; o 18º no jornal ZH; e o 19º no JM e no DG.

Após o **não**, as três próximas palavras duplicadas mais frequentes são as formas do verbo ser **é** e **foi**, seguidos da palavra **texto**. O *type* **texto** não consta na lista dos duplicados dos jornais FSP e ZH, apenas no JM, na segunda colocação, e no DG, na quarta posição. Como em nossas listas de *types*, constam todas as formas de uma mesma palavra, na lista de duplicados do DG e do CE, consta a palavra **danos** (plural), e **dano** (singular) consta somente nos *types* exclusivos do CE. Nas sentenças, das 1620 ocorrências de **dano**, em 960 faz parte da expressão terminológica **dano moral**. Quanto a **danos** (plural), das 2158 ocorrências, em 1379 compõe o plural de **dano moral: danos morais**. Em 299 compõe a expressão **danos materiais**.

Na coluna em azul, constam as palavras exclusivas do DG. Seguindo a tendência observada no jornal JM, a grande maioria ds *types* exclusivos do DG são palavras relacionadas ao esporte – **Grêmio**, **Inter**, etc. – e ao lazer – **show**, **ingressos**. As matérias são dos anos de 2008 e 2010, mas chama a atenção não haver, entre elas, alguma que se faça referência a um político daquele momento. Com o DG também parece se repetir o propósito editorial dos jornais populares: priorizar matérias de entretenimento em vez de matérias de cunho político e econômico. Mais um trecho de Amaral a confirmar esse posicionamento: “Os assuntos que interessam são prioritariamente os que mexem de

imediate com a vida da população. Na pauta, o atendimento do SUS e do INSS, a segurança pública, o mercado de trabalho, o futebol e a televisão” (AMARAL, 2006, p. 6).

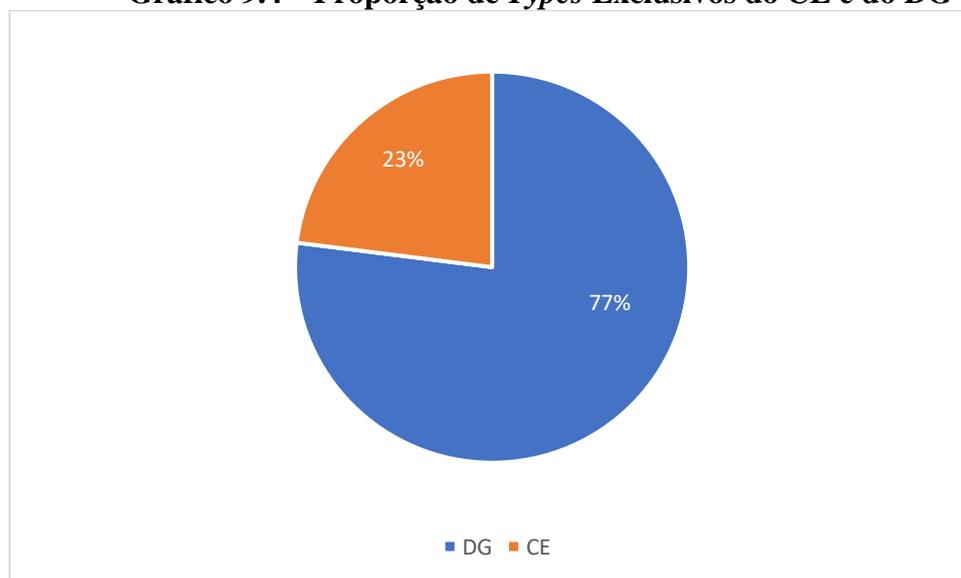
Tendo esse panorama inicial do léxico empregado no CE em comparação com o DG, passamos, agora, a analisar a relação entre os *types* exclusivos que se encontram no CE e no DG. A Tabela 9.14 apresenta os dados relativos a essa comparação.

Tabela 9.14 – *Types* Exclusivos do CE e do DG

<i>CORPUS</i>	Quantidade de <i>types</i> exclusivos	Percentual de <i>types</i> exclusivos
DG	22.255	77%
CE	6.797	23%
TOTAL	29.052	100%

Os percentuais colocados na última coluna da Tabela 9.14 são referentes à quantidade de *types* exclusivos de cada *corpus* em relação ao total de *types* exclusivos, conforme a última linha da segunda coluna (29.052). O Gráfico 9.4 ilustra esse cálculo.

Gráfico 9.4 – Proporção de *Types* Exclusivos do CE e do DG

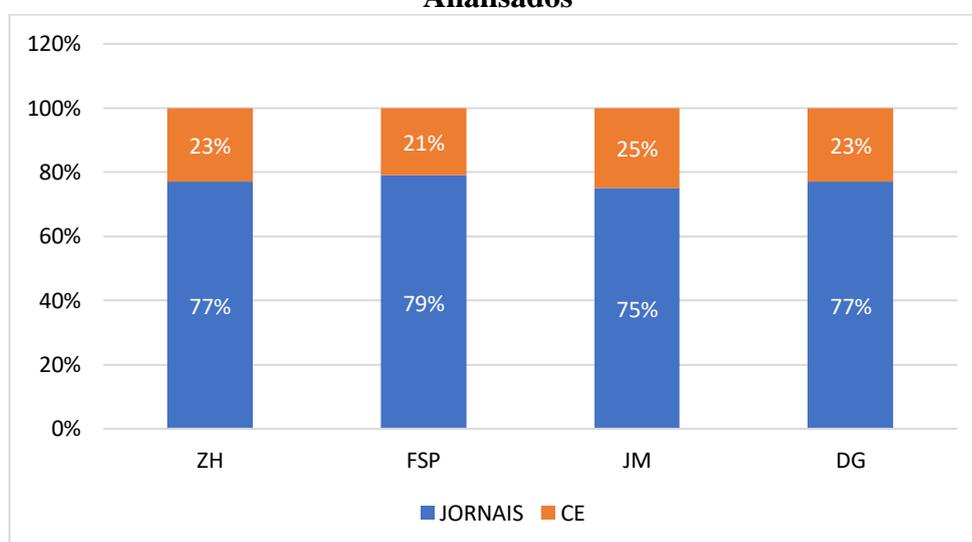


Pelo Gráfico 9.4, podemos entender que, do total de *types* exclusivos do CE e do DG juntos, 23% referem-se apenas a palavras que se encontram nas Sentenças, e 77% a palavras que se encontram apenas no DG. Considerando que nosso foco está nas Sentenças, podemos pressupor que **o leitor do DG terá, em tese, problemas em compreender 23% do vocabulário que se encontra nas Sentenças aqui analisadas.**

Entretanto, se relacionarmos o número de palavras exclusivas do CE (6.797), conforme Tabela 9.14, com o número total de *types* que este mesmo *corpus* apresenta (14.665), conforme Tabela 9.12, chegaremos a um percentual de 46,34%. Assim, com base nessa relação, **é também possível pressupor que o leitor do DG terá, em princípio, chance de não compreender aproximadamente 46% do léxico das sentenças aqui analisadas.**

O Gráfico 9.5, abaixo, sintetiza os dados dos *types* exclusivos de todos os jornais aqui analisados em relação aos *types* exclusivos do CE.

Gráfico 9.5 – Percentual de *Types* Exclusivos no CE e nos Jornais Analisados



Pelo Gráfico 9.5, é possível observar que o percentual médio de *types* exclusivos do CE em relação a todos os jornais juntos ficou em 23%. Em relação ao JM é que o CE ficou com um percentual maior de *types* exclusivos – 25%.

Se considerarmos os jornais do CCA – ZH e FSP -, o percentual médio de *types* exclusivos do CE fica em 22%. Se considerarmos os jornais do CCB – JM e DG -, o percentual médio de *types* exclusivos do CE fica em 24%.

Tendo em conta esses percentuais e as observações que fizemos ao longo das análises, o leitor dos jornais do grupo CCA terá, em princípio, dificuldade em compreender 22% do vocabulário empregado nas sentenças aqui analisadas. E o leitor dos jornais do grupo CCB terá, em tese, dificuldade de compreender 24% do vocabulário empregado nas sentenças aqui analisadas.

Outra comparação muito importante que fizemos ao longo de nossas análises foi quanto ao número de *types* exclusivos do CE resultante da comparação com aquele *corpus* de jornal em relação ao seu total de *types* exclusivos. A Tabela 9.15 traz esses dados novamente

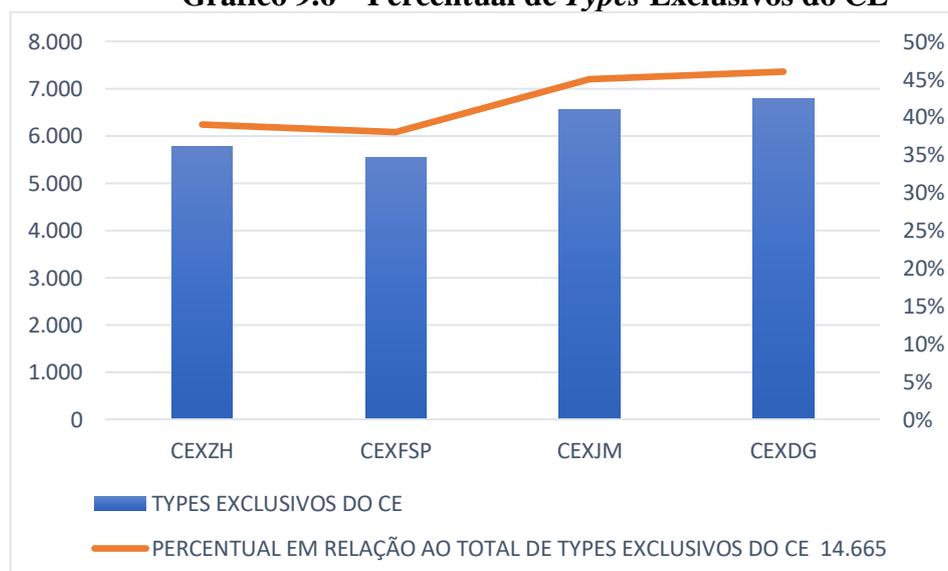
Tabela 9.15 – Percentual de *Types* Exclusivos do CE

	TYPES EXCLUSIVOS DO CE	PERCENTUAL EM RELAÇÃO AO TOTAL DE TYPES EXCLUSIVOS DO CE (14.665)
CEXZH	5.774	39%
CEXFSP	5.545	38%
CEXJM	6.559	45%
CEXDG	6.797	46%

Na segunda coluna da Tabela 9.15, consta o número de *types* exclusivos do CE encontrados na comparação do jornal relacionado na primeira coluna. Assim, o número 5.774 indica a quantidade de *types* exclusivos do CE na comparação com o jornal ZH; o número 5.545 indica a quantidade de *types* exclusivos do CE na comparação com o jornal FSP, e assim o número 6.559 com o JM, e o número 6.797 com o DG.

A última coluna traz o percentual encontrado com o cálculo da relação de cada número da segunda coluna com o total de *types* exclusivos do CE (14.665). O Gráfico 9.6 ilustra essa relação.

Gráfico 9.6 – Percentual de *Types* Exclusivos do CE



Os percentuais da Tabela 9.15 e ilustrados no Gráfico 9.6 indicam que os leitores dos jornais analisados, em princípio, não conhecem essas palavras do CE, visto que elas se encontram apenas nas sentenças e não em seus textos. Disso, então, se pode pressupor que, em relação ao vocabulário empregado nas sentenças aqui analisadas,

- a) o leitor do jornal ZH terá, em tese, 39% de dificuldade de compreensão;
- b) o leitor do jornal FSP terá, em tese, 38% de dificuldade de compreensão;
- c) o leitor do jornal JM terá, em tese, 45% de dificuldade de compreensão; e
- d) o leitor do jornal DG terá, em tese, 46 % de dificuldade de compreensão.

Ao finalizarmos a apresentação do *corpus* de contraste CCA – ZH e FSP -, salientamos o quão complexos os textos do CE eram para os leitores desses jornais. Se já eram complexos para leitores representados em grande parte por brasileiros com nível superior completo ou inacabado, como serão para os leitores do *corpus* de contraste CCB – os jornais populares DG e JM?

Os índices das letras *c* e *d* anteriores já respondem: o percentual de *types* do CE ausentes nos textos desses jornais é bem maior: 45% e 46%. E esse resultado, conforme já comentamos, é previsível, uma vez que o leitor-alvo desses jornais são pessoas das classes C, D e E, que normalmente têm menos escolaridade e experiência em leitura que as classes A e B.

Cabe aqui retomar algumas observações colocadas ao longo deste trabalho. Uma delas é a observação feita na seção 6.3 quando analisamos os dados do questionário que realizamos. Julgamos pertinente, inclusive, repetir a citação lá colocada:

Estudos de Hu e Nation (*apud* Nation, 2001: 147) evidenciaram uma nítida relação entre compreensão de texto e de vocabulário, verificando que a compreensão do texto melhora na medida em que aumenta o número de palavras conhecidas. Se o texto apresenta 95% de palavras conhecidas, alguns leitores compreendem o texto, mas muitos não adquirem adequada compreensão; se a porcentagem é de 90% de palavras conhecidas, somente um pequeno grupo de pessoas consegue compreender o texto; e no nível de 80% de vocabulário conhecido, nenhum leitor consegue compreender o texto. Hu e Nation concluíram que, **para garantir a compreensão, um texto deve conter por volta de 98% de palavras conhecidas.** Carver (1994: 432, *apud* Nation, 2001: 148) acrescenta ainda que **um texto apropriado, ou seja, não considerado difícil, contém somente 1% ou menos de palavras desconhecidas.** (FULGÊNCIO E LIBERATO, 2010, p. 105 – grifos nossos)

Tendo em conta a citação acima, percentuais de *types* desconhecidos iguais ou superiores a 38%, como ocorre com as Sentenças aqui analisadas, tornam tais textos

ininteligíveis para os leitores dos jornais aqui analisados, independentemente de serem de referência ou populares.

É preciso ponderar aqui que o CE é composto por textos especializados que veiculam uma linguagem também especializada. Isso pressupõe o emprego de muita terminologia que, evidentemente, pode ser desconhecida de um leigo na área. E a presença de terminologia também aumenta o número de palavras abstratas no CE, aumentando a complexidade do texto.

A maior presença de palavras abstratas no CE se evidencia nas listas comparativas das Tabelas 9.4; 9.7; 9.10 e 9.13. É mais difícil elaborar uma imagem mental para **demandada, inversão, homologatória, recursais**, do que para **Bolsonaro, Inter, Grêmio, diretor, chefe**, etc.

Mesmo que não se possa fugir do emprego da terminologia própria da área, é preciso atentar para o fato de que os textos do CE são Sentenças dos JECs. Esses Juizados, como já referimos, foram criados para promover maior acesso à Justiça e atendem casos que têm como valor-limite 40 salários mínimos. Além disso, o destinatário dessas Sentenças não é um especialista, mas o cidadão que busca a Justiça. Ainda que possa haver a intermediação de um advogado em casos que vão de 20 a 40 salários mínimos, o destinatário final das decisões judiciais dos JECs, assim como a de qualquer outra decisão judicial, é sempre o cidadão comum, normalmente leigo em Direito.

Por isso, estratégias de simplificação lexical, com empregos de sinônimos mais simples, ou por explicação como apresentamos no Estudo-Piloto 2, são recursos que poderiam ser empregados para facilitar a compreensão do vocabulário desses textos. Infelizmente não é isso o que acontece, como podemos ver com a expressão ***inversão do ônus da prova***, nos exemplos a seguir.

(9.3) **INVERSÃO DO ÔNUS DA PROVA.** A ***inversão do ônus da prova***, prevista no art. 6º, VIII, do CDC, constitui faculdade do magistrado, que, verificando a existência de verossimilhança nas alegações da parte autora ou a sua hipossuficiência perante a parte demandada, facilitará a defesa de seus direitos.

(9.4) Neste sentido, a ***inversão do ônus da prova ope judicis*** é imperiosa, eis que presentes os seus requisitos, quais sejam, a verossimilhança das alegações da parte autora, bem como, (sic) sua hipossuficiência ante ao (sic) réu, sendo certo que tais requisitos são alternativos e não cumulativos.

Para melhor compreensão do que significa **inversão do ônus da prova**, deve-se ter em mente a seguinte regra: quem alega um fato deve provar este fato. Este é o **ônus** (obrigação) de quem alega.

Nos exemplos 9.3 e 9.4, a preocupação foi em esclarecer quais os requisitos (exigências) para se adotar a inversão do ônus da prova: **verossimilhança** e **hipossuficiência**. E as palavras referentes a esses requisitos não constam em nenhum dos *corpora* de jornais aqui analisados. Em um dos exemplos, ocorre, inclusive, uma expressão em latim – *ope judicis*. Ou seja, os trechos partem do pressuposto de que as expressões utilizadas são do conhecimento do seu leitor. E isso, como atestam os dados aqui apresentados, não é verdadeiro, pelo menos quanto aos leitores dos jornais dos *corpora* CCA e CCB.

Em vez de **verossimilhança**, por exemplo, poderia ser utilizada a expressão um pouco mais simples **aparência de verdade**, e, em vez de **hipossuficiência**, poderia se colocar **inferioridade econômica**. Quanto à última sugestão, verificamos que **inferioridade** não se encontra nos textos de jornais aqui analisados. Por isso, sugerimos **desvantagem econômica**. **Desvantagem** pode ser considerada um sinônimo para **inferioridade** e consta, inclusive, nos jornais DG e JM.

A seguir apresentamos uma sugestão de redação simplificada para o exemplo 9.4.

Reescrita 1 do Exemplo 9.4

*Neste sentido, a inversão do ônus da prova (a substituição da parte que tem o dever de provar o que é alegado), de acordo com a decisão judicial, é necessária. E assim é porque a parte autora se encontra em **desvantagem econômica** em relação ao réu e as suas alegações parecem ser verdadeiras. Essas exigências podem aparecer **em conjunto** ou **em separado**.*

No Quadro 9.4, apresentamos os itens do trecho original e as sugestões de simplificação de modo individualizado.

Quadro 9.4 – Sugestão de Simplificação de Modo Individualizado

Trecho Original	Alterações realizadas
(1) a inversão do ônus da prova	(1) (a substituição da parte que tem o dever de provar o que é alegado)
(2) <i>ope judicis</i>	(2) de acordo com a decisão judicial,
(3) é imperiosa	(3) é obrigatória

[5] eis que	[5] E assim é ... Iniciada nova frase.
(6) a verossimilhança das alegações da parte autora, (7) hipossuficiência	(6) parecem ser verdadeiras (7) desvantagem econômica
[8] sendo certo que tais (4) requisitos	[8] Essas Iniciada nova frase. (4) exigências
(8) são alternativos e não cumulativos	(8) podem aparecer em conjunto ou em separado.

Na reescrita apresentada, utilizamos alguns procedimentos simplificadoros, como a inserção de uma explicação entre parênteses, como ocorreu com o item (1) do Quadro 9.4 e a substituição por sinônimos mais simples, como ocorreu com os itens (2), (3), (4), (6), (7) e (8). Também reduzimos a extensão do período, criando duas frases: [5] e [8].

E, considerando o perfil de leitor que temos, um adulto com pouca escolaridade e experiência em leitura, é possível, ainda, simplificar um pouco mais o trecho. É o que se vê a seguir.

Reescrita 2 do Exemplo 9.4

A pessoa que acusa outra em um processo judicial tem sempre o dever de provar a verdade do que diz. Mas esse dever de provar pode ser alterado pelo juiz quando ele considerar que: a) quem acusa tem menos condições econômicas do que o acusado; e b) que o que diz parece ser verdadeiro. Esses dois elementos a) e b) podem ser avaliados pelo juiz juntos ou em separado para alterar a obrigação do acusador de ter provar o que diz (chamado de inversão do ônus da prova).

No próximo capítulo, em que apresentamos algumas sugestões de simplificação, detalharemos esses procedimentos.

A partir daqui, seguindo nas comparações do CE com os *corpora* de contraste, vejamos o que ocorre nas relações entre o CE e o *Corpus* de Contraste CCP, terceiro *corpus* de contraste.

9.3 CE X CORPOP

O *Corpus* de Contraste CCP - CorPop (PASQUALINI, 2018), conforme mencionamos na seção 5.1.4 deste trabalho foi compilado com base no nível de

letramento médio dos leitores do país e das características que poderiam compor um padrão de simplicidade textual em um *corpus* de textos adequados a esses leitores. Por se constituir desses materiais-fonte diferenciados, o CorPop extrapola o gênero jornalístico e serve para oferecer uma referência para português popular brasileiro escrito em uso pela maioria dos brasileiros.

Para a comparação com o CorPop, procedemos da mesma forma como fizemos com os jornais. Os dados dessa comparação constam na Tabela 9.16, a seguir.

Tabela 9.16 – Dados dos *Types* do CE e do CorPop

<i>Corpus</i>	Total de <i>types</i>	Quantidade de <i>types</i> duplicados		Quantidade de <i>types</i> exclusivos	
		Número bruto	Percentual	Número bruto	Percentual
CorPop	32.311	8.338	18%	23.973	51%
CE	14.665	8.341	18%	6324	13%
TOTAL	46.976	16.679	36%	30297	64%

Os percentuais colocados na Tabela 9.16 referem-se à relação entre o número bruto de *types* (duplicados e exclusivos) e o total de *types* de ambos os *corpora*. Os dados da Tabela 9.16, acima, podem ser melhor visualizados na Figura 9.6, a seguir.

Figura 9.6 – Percentuais de *Types* Exclusivos e Duplicados do CE e do CorPop



Para exemplificarmos alguns casos dessas ocorrências, criamos a Tabela 9.17, a seguir, que traz as 20 palavras lexicais mais frequentes de cada conjunto.

Tabela 9.17 – Distribuição dos 20 *Types* mais frequentes na Comparação entre o CE e o CorPop

EXCLUSIVO CE (13%)		DUPLICADOS (36%)		EXCLUSIVO CORPOP (51%)	
FREQ.	TYPE	FREQ.	TYPE	FREQ.	TYPE
791	demandada	16501	não	815	Gabriela
773	ônus	3978	mais	708	amor
750	recursais	2767	autora	526	bala
564	opino	2628	lei	512	Renato
514	demandante	2378	art.	512	vou
434	CDC (Código de Defesa do Consumidor)*	2181	danos	453	homens
425	improcedência	2506	parte	416	vamos
422	homologatória	3695	ser	414	cabeça
413	homologo	2048	ré	396	Petrus
397	intimadas	1921	sentença	368	voz
393	leigo	1927	autor	362	medo
388	produza	1878	valor	361	coração
372	<i>quantum</i>	1693	nº	359	sei
370	inominado	2051	caso	353	quero
352	inexistência	1947	tem	344	tu
349	cancelamento	1631	dano	343	padre
327	inciso	1603	disse	332	pernas
310	faturas	1499	morais	322	Peri
306	dispensado	1489	recurso	319	vocês
296	citação	1440	autos	318	Gioconda

Na coluna em laranja, constam as primeiras 20 palavras exclusivas mais frequentes do CE. Aqui, o primeiro *type* é **demandada**. Com exceção de 4 palavras – **ônus**, **cancelamento**, **dispensado** e **citação** –, as demais palavras exclusivas do CE nessa comparação também o foram na comparação com os jornais.

Na coluna em verde, estão os *types* duplicados. O primeiro *type* mais frequente entre os duplicados continua sendo o **não**. Com exceção das formas verbais **tem** e **disse**, as demais palavras desta lista também apareceram na comparação com os jornais.

Na coluna em azul, constam as palavras exclusivas do CorPop. Chama a atenção a presença de 5 nomes próprios, inclusive a palavra mais frequente é um deles: **Gabriela**. Tais nomes referem-se a personagens de textos de literatura que fazem parte do CorPop. E a grande maioria dos demais *types* também parece fazer parte deste universo literário. É o caso de:

- a) **amor** (segunda mais frequente), **coração**, **medo**;
- b) formas verbais na primeira pessoa: **vou**, **sei**, **quero**;
- c) pronomes pessoais retos: **tu**, **vocês**

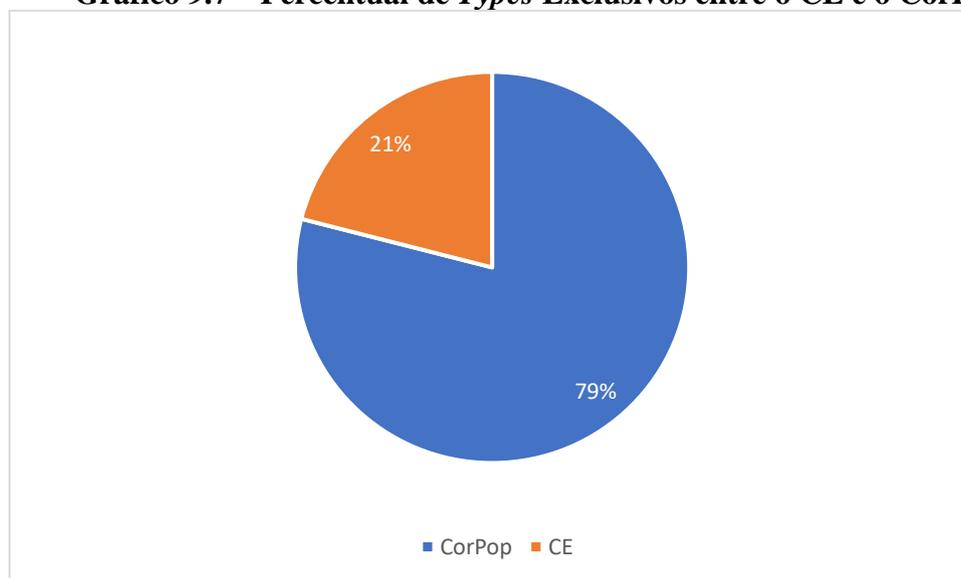
Tendo esse panorama inicial do léxico empregado no CE em comparação com o CorPop, passamos, agora, a analisar a relação entre os *types* exclusivos que se encontram no CE e no CorPop. A Tabela 9.18 apresenta os dados relativos a essa comparação.

Tabela 9.18 – *Types* Exclusivos do CE e do CorPop

<i>CORPUS</i>	Quantidade de <i>types</i> exclusivos	Percentual de <i>types</i> exclusivos
CorPop	23.973	79%
CE	6.324	21%
TOTAL	30297	100%

Os percentuais colocados na última coluna da Tabela 9.18 são referentes à quantidade de *types* exclusivos de cada *corpus* em relação ao total de *types* exclusivos, conforme a última linha da segunda coluna (30.297). O Gráfico 9.7 ilustra esse cálculo.

Gráfico 9.7 – Percentual de *Types* Exclusivos entre o CE e o CorPop



Pelo Gráfico 9.7, podemos entender que, do total de *types* exclusivos do CE e do CorPop juntos, 21% referem-se apenas a palavras que se encontram nas Sentenças, e 79% a palavras que se encontram apenas no CorPop. Considerando que os textos integrantes

do CorPop foram selecionados com base no nível de letramento médio dos leitores do país, é possível pressupor que **uma pessoa com nível de letramento médio terá, em tese, problemas em compreender 21% do vocabulário que se encontra nas Sentenças aqui analisadas.**

Entretanto, se relacionarmos o número de palavras exclusivas do CE (6.324), conforme Tabela 9.18, com o número total de *types* que este mesmo *corpus* apresenta (14.665), conforme Tabela 9.16, chegaremos a um percentual de 43,12%. Assim, com base nessa relação, **é também possível pressupor que cidadão brasileiro com nível de letramento médio terá, em princípio, chance de não compreender aproximadamente 43% do léxico das sentenças aqui analisadas.**

Comparando esse percentual de 43% com os percentuais dessa relação encontrados nos *corpora* de contraste CCA e CCB, temos a Tabela 9.19, a seguir.

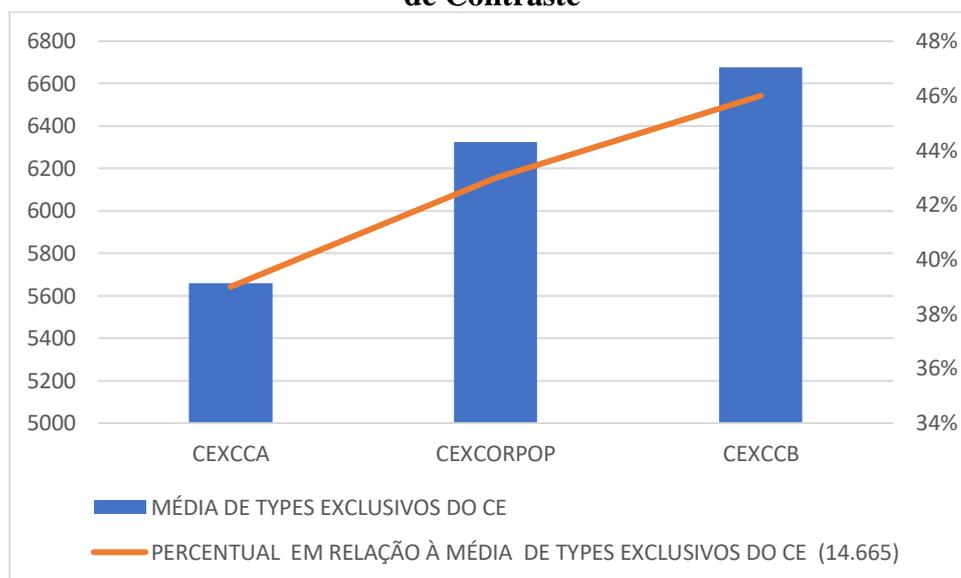
Tabela 9.19 – Percentual de *Types* Exclusivos do CE em relação aos *Corpora* de Contraste

	MÉDIA DE <i>TYPES</i> EXCLUSIVOS DO CE	PERCENTUAL EM RELAÇÃO À MÉDIA DE <i>TYPES</i> EXCLUSIVOS DO CE (14.665)
CEXCCA	5659	39%
CEXCCB	6678	46%
CEXCORPOP	6324	43%

A média apresentada na segunda coluna da Tabela 9.19 em relação ao CCA e ao CCB foi calculada com base no total de *types* exclusivos do CE em relação a cada jornal desses *corpora*. O número relativo ao CorPop traz o número total de *types* exclusivos do CE em relação a este *corpus*. Os percentuais da última coluna foram calculados com base na relação dos números da segunda coluna com o total de *types* do CE (14.665).

Os dados dessa Tabela 9.19 geraram o Gráfico 9.8

Gráfico 9.8 – Percentual de *Types* Exclusivos do CE em relação aos *Corpora* de Contraste



O eixo vertical esquerdo do Gráfico 9.8 representa o número de *types* exclusivos do CE, e o eixo vertical direito, o percentual que esses números representam. As colunas representam os *corpora* contrastados.

O Gráfico 9.8 permite visualizar com mais clareza que o léxico empregado nas Sentenças aqui analisadas está mais distanciado dos jornais do CCB (Diário Gaúcho e Jornal Massa). Em segundo nesse *ranking* encontra-se o CorPop, e por último os jornais do CCA (Folha de São Paulo e Jornal Zero Hora).

Os dados da comparação do CE em relação ao CorPop confirmam que ele representa o nível de letramento médio dos leitores do país, porque ficou exatamente entre os jornais de referência e os jornais populares.

Além disso, o percentual de 43% de *types* do CE ausentes no CorPop é mais um dado a confirmar a distância que o vocabulário empregado nas Sentenças mantém do cidadão brasileiro.

Tendo exposto as comparações entre o CE os *corpora* de contraste, passamos a verificar como se dá a relação do nosso CE com os *corpora* de referência.

9.4 CE X *CORPORA* DE REFERÊNCIA (CR1 E CR2)

O *Corpus* de Referência 1 (CR1) é composto por obras literárias brasileiras publicadas entre os anos de 1840 a 1956, que, em princípio, apresentam uma linguagem clássica mais antiga.

Adotando inicialmente o procedimento comparativo realizado com os *corpora* de contraste com o CR1, obtivemos os dados constantes na Tabela 9.20, a seguir.

Tabela 9.20 – Dados dos *Types* do CE e do CR1

<i>Corpus</i>	Total de <i>types</i>	Quantidade de <i>types</i> duplicados		Quantidade de <i>types</i> exclusivos	
		Número bruto	Percentual	Número bruto	Percentual
CR1	82.136	9.720	10%	72.416	75%
CE	14665	9.724	10%	4.941	5%
TOTAL	96.801	19.444	20%	77.357	80%

Os percentuais colocados na Tabela 9.20 referem-se à relação entre o número bruto de *types* (duplicados e exclusivos) e o total de *types* (96.801) de ambos os *corpora*. Os dados da Tabela 9.20, acima, podem ser melhor visualizados na Figura 9.7, a seguir.

Figura 9.7 – Percentuais de *Types* Exclusivos e Duplicados do CE e do CR1



Para exemplificarmos alguns casos dessas ocorrências, criamos a Tabela 9.21, a seguir, que traz as 20 palavras lexicais mais frequentes de cada conjunto.

Tabela 9.21 – Distribuição dos 20 *Types* mais frequentes na Comparação entre o CE e o CR1

EXCLUSIVO CE (5%)		DUPLICADOS (20%)		EXCLUSIVO CR1 (75%)	
FREQ.	TYPE	FREQ.	TYPE	FREQ.	TYPE
791	demandada	39720	não	2017	coração
750	recursais	12300	mais	1930	amor

564	opino	7133	foi	1786	moça
538	recursal	5517	ser	1615	cabeça
514	demandante	5381	ainda	1553	velho
450	CPC (Código de Processo Civil)*	4870	tinha	1533	voz
434	CDC (Código de Defesa do Consumidor)*	4638	bem	1372	padre
430	indevida	4535	disse	1282	sei
422	homologatória	4499	há	1277	homens
413	homologo	4282	casa	1031	rosto
397	intimadas	4235	depois	997	idéia
388	produza	4193	só	940	dizia
370	inominado	3716	olhos	917	menina
352	inexistência	3629	estava	891	moço
349	cancelamento	3503	dia	880	rei
349	contratual	3348	tempo	862	sol
327	inciso	3031	havia	860	jorge
327	juizado	2955	homem	851	ouro
316	IGP (IGP-M – Índice Geral de Preços do Mercado)*	2875	senhor	831	quero
295	alega	2788	autora	818	ah

Na coluna em laranja, constam as primeiras 20 palavras exclusivas mais frequentes do CE. Aqui, o primeiro *type* é **demandada**, tal qual ocorreu com o CorPop, com o JM e com o jornal ZH. Com exceção de 2 palavras – **contratual** e **alega** -, as demais palavras exclusivas do CE nessa comparação também o foram na comparação com os jornais.

Na coluna em verde, estão os *types* duplicados. O primeiro *type* mais frequente entre os duplicados continua sendo o **não**. Dos elementos duplicados além do **não**, apenas 5 também se encontram nas tabelas relativas aos jornais e ao CorPop: **mais**, **foi**, **ser**, **disse** e **autora**.

Na coluna em azul, constam as palavras exclusivas do CR1. As duas primeiras mais frequentes – **coração** e **amor** – refletem a temática quase sempre frequente na literatura: o romance. Observa-se a grafia de **idéia** com acento, o que indica também a época das obras, antes da última reforma ortográfica. E chama a atenção a presença da interjeição **ah**, com 818 ocorrências.

Antes de prosseguirmos, cumpre referir que, tanto na Tabela 9.20 quanto na Figura 9.7, chama a atenção o percentual de *types* exclusivos do CE em 5%. É preciso ponderar, porém, que aqui a relação se estabelece com um conjunto maior de *types* (82.136).

A tabela 9.22, a seguir, traz os dados dos *types* exclusivos do CE em relação a todos os *corpora* aqui analisados.

Tabela 9.22 – *Types* Exclusivos do CE

<i>Corpus em comparação com o CE</i>	Total de <i>types</i>	Quantidade de <i>types</i> exclusivos do CE	
		Número bruto	Percentual
ZH+CE	43.295	5.774	13%
CE+FSP	44.998	5.545	12%
CE+JM	42.655	6.559	15%
CE+DG	44.861	6.797	15%
CE+CORPOP	46.976	6.324	13%
CE+CR1	96.801	4.941	5%

A segunda coluna da Tabela 9.22 apresenta a soma de *types* do CE com os *types* de cada *corpus* analisado. Por exemplo, os 43.295 *types* da linha relativa à comparação entre o CE e o jornal ZH representam a soma de 14.665 (CE) + 28.630 (ZH). E o percentual colocado na última coluna da tabela são resultado da relação do número bruto de *types* exclusivos do CE em relação ao total de *types* da segunda coluna. Assim, os 13% da primeira linha é resultado da relação entre 5.774 e 43.295.

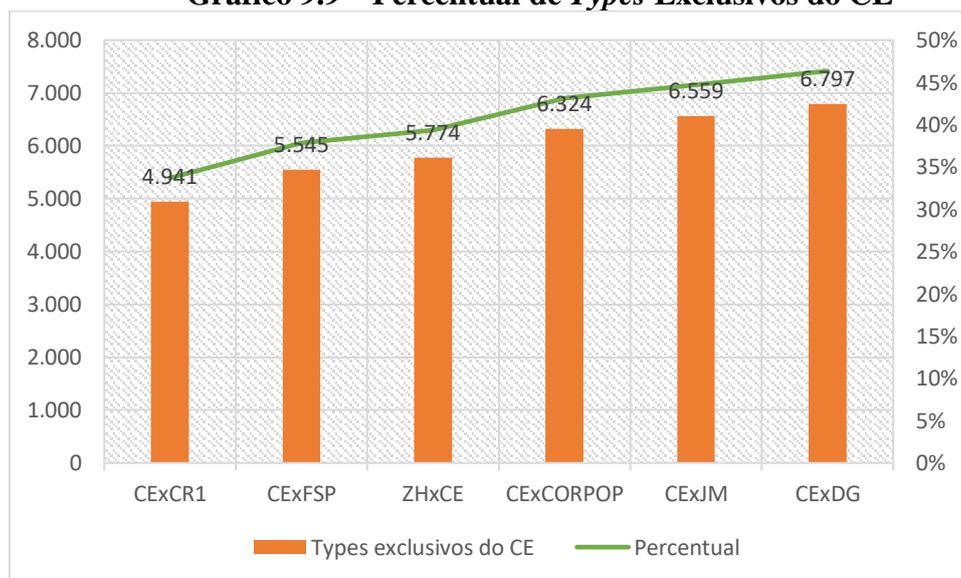
A soma de *types* do CR1 com o CE é um pouco mais que o dobro que as somas com os outros *corpora*, e o número de *types* exclusivos do CE é um pouco menor que os *types* exclusivos em relação aos outros *corpora*. E o percentual ficou em torno de 50% a menos que os demais percentuais. Assim ocorre porque a relação se estabelece com um conjunto maior de *types* (96.801).

Desse modo, julgamos que a relação entre o número de *types* exclusivos do CE resultante da comparação com cada *corpus* aqui analisado (números da coluna *número bruto* da Tabela 9.22) em relação ao seu total de *types* é que mais facilmente dimensiona as particularidades lexicais do CE. A Tabela 9.23 traz esses dados.

Tabela 9.23 – Percentuais de *Types* Exclusivos do CE

<i>Corpus</i> em comparação com o CE	<i>Types</i> exclusivos do CE	Total de <i>types</i> do CE	Percentual
CExCR1	4.941	14.665	34%
CExFSP	5545	14.665	38%
ZHxCE	5774	14.665	39%
CExCORPOP	6324	14.665	43%
CExJM	6.559	14.665	45%
CExDG	6.797	14.665	46%

O percentual colocado na última coluna da Tabela 9.23 é resultado da relação entre o número de *types* exclusivos do CE de cada conjunto e o total de *types* do CE. Assim, por exemplo, o percentual de 34% da primeira linha (CEXCR1) é resultado da relação entre 4.941 e 14.665. Os dados da Tabela 9.23 estão representados no Gráfico 9.9.

Gráfico 9.9 – Percentual de *Types* Exclusivos do CE

Os números colocados no eixo vertical esquerdo do Gráfico 9.9 indicam o número de *types* exclusivos do CE (apontados detalhadamente acima de cada coluna laranja). Os percentuais indicados no eixo vertical direito indicam os percentuais que cada um desses números representa em relação ao total de *types* do CE (14.665). Cada coluna em laranja representa uma relação do CE com os *corpora* aqui analisados.

Com os dados da Tabela 9.23 e do Gráfico 9.9, verificamos que o menor percentual de palavras exclusivas do CE é de 34%, relativo à comparação com o CR1. E o maior é de 46%, relativo à comparação com o DG. Com isso, podemos concluir que os

padrões lexicais das sentenças aqui analisadas estão mais próximos do CR1 e mais distantes do DG.

Entre essas extremidades, o segundo *corpus* com o padrão lexical mais próximo do CE é o jornal FSP, seguido pelo jornal ZH. Em quarto lugar, está o CorPop e, em quinto o JM.

Comentaremos mais sobre esses resultados após as comparações relativas às palavras-chave e multipalavras do CE e dos *corpora* de referência – CR1 e CR2 – que passamos a apresentar.

9.4.1 CE X CR1 – PALAVRAS-CHAVE

Ainda no que se refere ao CR1, por ser ele de referência, será utilizado também para extrairmos as palavras-chave do CE. E a extração das palavras-chaves é um processo importante porque, pela análise contrastiva entre os dois *corpora*, é possível verificar as palavras que se destacam pela frequência no *corpus* de estudo. Fromm (2008, p. 12) explica que, se a frequência de uma palavra é mais proeminente no *corpus* de estudo do que no *corpus* de referência, o sistema a coloca na lista de palavras-chave.

Como vimos na Tabela 9.1, o CR1 contém 2.460.268 *tokens*, e o CE 542.836. Ou seja, o CR1 é quase 5 vezes maior que o CE em relação ao número total de palavras. Essa proporção atende à recomendação proposta por Berber Sardinha (2004, p. 102):

Os tamanhos críticos de *corpora* de referência são 2, 3 e 5 vezes o tamanho do *corpus* de estudo. *Corpora* de referência com essas dimensões retornam significativamente mais palavras-chave do que *corpora* de tamanhos menores.

Além disso, importa observar que o CR1 é composto por textos da literatura brasileira e, em princípio, não abarcam o nosso CE. Ou seja, são conjuntos de textos de gêneros diferentes. Essa diferença influencia nos tipos de palavras que podem se tornar chave:

Ao comparar um *corpus* de estudo de artigos de pesquisa acadêmicos de medicina com um *corpus* de referência do mesmo tipo, pode-se esperar que palavras como *resultados*, *análise* e *sugerem* não se tornem chave. Já um *corpus* de referência de um gênero distinto do de estudo tende a não excluir tais palavras *genéricas*. Por isso, um *corpus* de referência geral, que inclua vários gêneros, é tido como a escolha não-marcada para estudos de palavras-chave”. (SARDINHA, 2004, p. 101)

Para o processamento das palavras-chave do CR1 e do CR2, valemo-nos da ferramenta Sketch Engine¹¹⁰, pois o CR2 está disponível somente nessa plataforma. Assim procedemos, para que os padrões de extração das palavras-chave ficassem uniformes nas comparações com ambos os *corpora* de referência.

A Tabela 9.24 traz os 20 primeiros itens das palavras-chave extraídas da comparação do CE com o CR1.

Tabela 9.24 – Vinte Primeiras Palavras-Chave do CE x CR1

Item	Frequência (CE)	Frequência (CR1)	Frequência Relativa (CE)	Frequência Relativa (CR1)	Pontuação (Chavicidade)
cível	1405	1	2161,1228	0,34595	1606,396
demandada	791	0	1216,68909	0	1217,689
recursais	750	0	1153,62427	0	1154,624
danos	2158	8	3319,36182	2,76757	881,3
opino	564	0	867,52545	0	868,525
débito	756	1	1162,85327	0,34595	864,71
indenização	1299	4	1998,07739	1,38379	838,614
recursal	538	0	827,53314	0	828,533
requerida	1092	3	1679,677	1,03784	824,735
demandante	513	0	789,07904	0	790,079
CDC (Código de Defesa do Consumidor)	433	0	666,02576	0	667,026
indevida	430	0	661,41125	0	662,411
homologatória	422	0	649,10596	0	650,106
consumidor	1291	6	1985,77209	2,07568	645,962
homologo	413	0	635,26245	0	636,262
relator	546	1	839,8385	0,34595	624,719
intimadas	397	0	610,65179	0	611,652
produza	388	0	596,80829	0	597,808
prestação	515	1	792,15533	0,34595	589,292
consideram-se	374	0	575,27399	0	576,274

Fonte: Sketch Engine¹¹¹

¹¹⁰ O Sketch Engine permite o *download* de no máximo 1.000 itens dessas listas.

¹¹¹ Traduzimos os títulos das colunas para o português.

A primeira coluna da Tabela 9.24 contém a lista das palavras que se destacaram na comparação entre o CE e o CR1. São consideradas palavras-chave e prováveis candidatas a termos no âmbito dos textos das sentenças aqui analisadas. A segunda e a terceira colunas trazem o número de frequência absoluta de cada palavra da primeira coluna nos *corpora* em comparação. A terceira e a quarta colunas trazem o número de frequência relativa, e a última seria a chavicidade. A chavicidade significa

o quanto a palavra em destaque, na relação entre o *corpus* de análise e o *corpus* de referência, é representativa na frequência relativa (o programa compara, estatisticamente, a frequência desta palavra em ambos os *corpora*; se ela apresenta um uso mais [ou menos] destacado no *corpus* de análise do que no de referência, ela é incluída na lista). (FROMM, 2008, p. 12)

Das palavras-chave listadas na Tabela 9.24:

- a) Doze não constam no CR1: **demandada, recursais, opino, recursal, demandante, CDC, indevida, homologatória, homologado, intimadas, produza e considerem-se**. Essas palavras também não constam em nenhum dos corpora aqui analisados.
- b) Sete não se encontram nas listas dos *types* exclusivos do CE relacionados nas Tabelas 9.4, 9.7, 9.10, 9.13, 9.17 e 9.21, referentes às comparações com os *corpora* anteriores: **danos, débito, indenização, consumidor, relator e prestação**.

Das palavras relacionadas no item **a**, temos 4 formas verbais¹¹² – **opino, homologado, produza e considerem-se**. Relembramos aqui a observação que fizemos no início de que as comparações com que trabalhamos são com as **formas** das palavras, e não com os **lemas**. Por isso, verificamos em cada *corpus* até aqui analisados se tais verbos poderiam neles constar conjugados em outros tempos, pessoas e modos.

Essa verificação se deu com o auxílio da ferramenta Sketch Engine, que possibilita pesquisar as palavras por categoria gramatical, na funcionalidade Lista de Palavras. A lista dos verbos gerada por essa funcionalidade apresenta os verbos lematizados. É também possível verificar todas as formas em que aparecem, porém nossa intenção foi apenas verificar se tais verbos constariam nos *corpora* analisados, independentemente da sua forma.

¹¹² Demandada, no contexto das sentenças, refere-se à pessoa contra quem se propõe a ação. Consideramos como substantivo. Intimadas – consideramos adjetivo.

A Tabela 9.25 apresenta, então, os dados da frequência dos verbos **opinar**, **homologar**, **produzir** e **considerar** em nossos *corpora* de contraste e no CR1.

Tabela 9.25 – Frequência dos verbos *opinar*, *homologar*, *produzir* e *considerar* nos *Corpora* de Contraste e no CR1

	OPINAR	HOMOLOGAR	PRODUZIR	CONSIDERAR
FSP	5	4	116	284
ZH	11	8	107	237
JM	5	2	56	158
DG	8	3	32	76
CR1	24	1	286	350
CORPOP¹¹³	sim	não	sim	sim

Pelos dados trazidos na Tabela 9.25, podemos verificar que outras formas de tais verbos se encontram nos *corpora* aqui analisados. A exceção se dá com o verbo **homologar**, que não consta no CorPop. Porém, cabem ainda algumas observações.

As formas verbais **opino** e **homologo** estão flexionadas na 1ª pessoa. Já comentamos sobre a primeira - **opino** - quando analisamos as listas de palavras relativas à comparação entre o CE e o jornal ZH. Lá registramos que é a forma como normalmente o juiz leigo conclui o seu parecer (Sentença) sobre o caso em julgamento.

E a forma verbal **homologo** é empregada pelo juiz togado, no final da Sentença redigida pelo juiz leigo, como se vê nos exemplos abaixo.

(9.5) *Nos termos do art. 40 da Lei no 9.099/95, **homologo** a proposta de decisão, para que produza efeitos como sentença.*

(9.6) ***HOMOLOGO** a sugestão de decisão de procedência, em parte, do pedido, consoante art. 40 da Lei 9.099/95.*

A forma verbal **produza**, flexionada no presente do subjuntivo, como um comando, nas 388 ocorrências se dá no contexto do exemplo (9.5):

(9.5) *Nos termos do art. 40 da Lei no 9.099/95, **homologo** a proposta de decisão, para que **produza** efeitos como sentença.*

¹¹³ O CorPop não se encontra no Sketch Engine, por isso não contabilizamos a frequência, mas apenas a ocorrência.

E a forma verbal **consideram-se**, ainda que esteja flexionada no presente do indicativo, expressa um comando de um ato que será realizado num futuro já designado. O exemplo (9.7) ilustra essa situação.

(9.7) *As partes **consideram-se** intimadas a partir da publicação da decisão, caso tenha ocorrido no prazo assinado; do contrário, a intimação terá de ser formal.*

Tais comandos aparecem, na maioria das vezes, na forma mais direta, com os verbos no imperativo, como ocorre com as seguintes formas verbais: **intimem-se**, **dê-se vista às partes**, etc.

As demais palavras do item a - **demandada**, **recursais**, **recursal**, **demandante**, **indevida**, **homologatória**, **intimadas** -, por constarem exclusivamente no CE, além de serem fortes candidatas a termo¹¹⁴, são palavras, em princípio, desconhecidas e, portanto, complexas aos leitores dos demais *corpora* aqui analisados.

Em relação às palavras do item b - **danos**, **débito**, **indenização**, **consumidor**, **relator** e **prestação** -, elas constam nos outros *corpora* aqui analisados, assim como no CR1. Foram listadas entre as palavras-chaves porque a sua frequência no CE se destaca em relação à frequência no CR1. É o que ocorre, por exemplo, com **danos**: no CE, com 542.836 *tokens*, **danos** ocorre 2.158 vezes; e no CR1, com 2.460.268 *tokens*, **danos** ocorre apenas 8 vezes.

Feita a exposição das palavras-chave do CE em relação ao CR1, vejamos como elas se dão com a comparação com o segundo *corpus* de referência de nosso trabalho.

9.4.2 CE X CR2 – PALAVRAS-CHAVE

O CR2 é representado pelo *Corpus* ptTenTen: *Corpus* of the Portuguese Web. Conforme mencionamos na seção 5.1.6, o CR2 é composto por um conjunto de textos coletados da internet – 76% de textos do Português Brasileiro e 24% do Português Europeu (24%).

Este *corpus* está disponível para consulta no sistema Sketch Engine e, segundo informações dessa ferramenta, apresenta 4.622.750.491 *tokens* e 3.896.392.719 palavras. Salientamos que *tokens*, para esta ferramenta, não são apenas as palavras, mas outros caracteres, como os de pontuação. As palavras corresponderiam ao total de *tokens* como faz o AntConc, ainda que possa haver variação nessa contagem.

¹¹⁴ Nosso enfoque não é tratar do que seria ou não termo. Apenas fizemos a observação.

A Tabela 9.26 traz os 20 primeiros itens das palavras-chave extraídas da comparação do CE com o CR2.

Tabela 9.26 – Vinte Primeiras Palavras-Chave do CE x CR2

Item	Frequência (CE)	Frequência (CR2)	Frequência Relativa (CE)	Frequência Relativa (CR1)	Pontuação (Chavicidade)
opino	564	614	867,52545	0,13282	766,692
recursais	750	2976	1153,62427	0,64377	702,423
homologo	413	458	635,26245	0,09908	578,907
demandada	791	5112	1216,68909	1,10584	578,245
homologatória	422	642	649,10596	0,13888	570,83
intimadas	397	988	610,65179	0,21373	503,946
inominado	370	645	569,12134	0,13953	500,314
ré	2045	36652	3145,54883	7,92861	352,412
demandante	513	6282	789,07904	1,35893	334,931
ipsa	217	384	333,78195	0,08307	309,105
improcedência	425	5347	653,72046	1,15667	303,579
indenizatório	291	2535	447,60623	0,54837	289,727
recursal	538	9089	827,53314	1,96615	279,33
intimem-se	191	577	293,78964	0,12482	262,078
requerida	1092	25141	1679,677	5,43854	261,034
quantum	372	6959	572,19763	1,50538	228,787
cível	1405	43781	2161,1228	9,47077	206,491
publique-se	187	1858	287,63699	0,40193	205,886
consideram-se	374	8571	575,27399	1,85409	201,912
igp-m	312	6450	479,90771	1,39527	200,774

Fonte: Sketch Engine

A Tabela 9.26 está organizada da mesma forma que a Tabela 9.24. Na primeira coluna consta a lista das palavras que se destacaram na comparação entre o CE e o CR2. A segunda e a terceira colunas trazem o número de frequência absoluta de cada palavra da primeira coluna nos *corpora* em comparação. A terceira e a quarta colunas trazem o número de frequência relativa, e a última seria a chavicidade.

Das palavras-chave listadas na Tabela 9.26, todas também ocorrem no CR2, e apenas 4 não se encontram nas listas de palavras exclusivas do CE expostas até aqui: **ré**, **ipsa**, **intimem-se** e **publique-se**.

A palavra **ré** consta, ainda que com baixa frequência, em todos os *corpora* de contraste e no CR1: no jornal FSP, ocorre 5 vezes; no jornal ZH, 1; no JM, 1; no DG, 1; no CorPop, 1; e no CR1, 16 vezes.

A palavra *ipsa*, que ocorre somente no CE, faz parte da expressão em latim **in re ipsa**, que significa *da própria coisa, do próprio fato*, algo que é presumido, que não precisa de prova. Fizemos referência a essa expressão quando analisamos as listas de *types* exclusivos do CE na comparação com o jornal ZH.

Quanto às formas verbais **intimem-se** e **publique-se**, retomamos as observações quanto aos verbos presentes na lista de palavras-chave extraídas da comparação do CE com o CR1. Como esses verbos estão flexionados, verificamos nos *corpora* aqui analisados se eles não se encontravam em outros tempos, pessoas e modos. Valendo-nos do mesmo procedimento adotado com o CR1, encontramos os dados apresentados na Tabela 9.27.

Tabela 9.27 – Frequência dos verbos *intimar* e *publicar* nos *Corpora* de Contraste e nos *Corpora* de Referência

	INTIMAR	PUBLICAR
FSP	5	229
ZH	4	171
JM	4	83
DG	6	89
CORPOP¹¹⁵	sim	sim
CR1	86	162
CR2	19.805	1.105.563

Pelos dados trazidos na Tabela 9.27, podemos verificar que outras formas de tais verbos se encontram nos *corpora* aqui analisados.

O verbo **intimar** ocorre com bem menos frequência que o **publicar**. Entre as acepções trazidas pelo Dicionário Eletrônico Houaiss (2007) ao verbo **intimar**, consta a seguinte: “3 Rubrica: termo jurídico. fazer intimação a”. Para o verbo **publicar**, o mesmo dicionário não coloca acepção com rubrica jurídica. Ou seja, o verbo **intimar** está mais relacionado à área do Direito do que o verbo **publicar**. Daí a sua menor frequência nos *corpora* de contraste e de referência em relação ao verbo **publicar**.

As formas verbais **intimem-se** e **publique-se**, no imperativo afirmativo, expressam um comando que normalmente ocorre no final da sentença, como já

¹¹⁵ O CorPop não se encontra no Sketch Engine, por isso não contabilizamos a frequência, mas apenas a ocorrência.

comentamos em relação à forma verbal **consideram-se**. Abaixo, colocamos mais um exemplo dessas situações.

(9.7) Nos termos do art. 40 da Lei n. 9.099/95, **remetam-se** os autos ao Juiz de Direito Presidente deste Juizado Especial Cível para a devida homologação. **Registre-se, Publique-se e Intimem-se.**

Além das formas verbais **intimem-se** e **publique-se**, destacamos outras formas verbais – **remetam-se** e **registre-se** –, também flexionadas no imperativo afirmativo presentes neste trecho recorrente no final das sentenças analisadas.

Após as análises das listas resultantes das comparações até aqui realizadas separadamente, julgamos pertinente verificar que palavras das listas de *types* exclusivos do CE se repetiram em todas as listas. A Tabela 9.28 traz esses dados.

Tabela 9.28 – Types Exclusivos do CE em relação a todos os *corpora* analisados

	Type	Frequência No CE	Quantidade de <i>corpora</i> em que NÃO ocorrem
1.	demandada	791	8
2.	recursais	750	8
3.	demandante	514	8
4.	homologatória	422	8
5.	homologo	413	8
6.	intimadas	397	8

A comparação aqui foi feita com as listas resultantes das comparações com os *corpora* de contraste – ZH, FSP, JM, DG e CorPop – e com os *corpora* de referência CR1 e CR2. Em relação ao CR1, temos duas listas: uma resultante da comparação realizada nos mesmos moldes da que se deu com os *corpora* de contraste; e outra relativa às palavras-chave. Com o CR2, a comparação se deu apenas com as palavras-chave. Participaram dessa comparação, então, 8 listas: dos jornais ZH, FSP, JM e DG; do CorPop, as duas do CR1 e a do CR2.

Como se vê na Tabela 9.28, as 6 palavras que ocorrem somente no CE são, em ordem de frequência: **demandada**, **recursais**, **demandante**, **homologatória**, **homologo** e **intimadas**. É interessante observar que há 4 palavras que se relacionam por fazerem

parte da mesma família de palavras: **demandada** e **demandante**; **homologatória** e **homologo**.

Demandada e **demandante** são palavras derivadas do verbo **demandar** e, no âmbito dos textos do CE, referem-se às partes envolvidas numa **demandada**. Quanto ao verbo **demandar**, o Dicionário Houaiss apresenta 8 acepções, em 2 há a rubrica **termo jurídico**, como se vê no Quadro 9.5.

Quadro 9.5 – Acepções do Dicionário Houaiss para o verbete *demandar*

transitivo direto e intransitivo

7 Rubrica: termo jurídico.

instaurar processo judicial (contra) pessoa, instituição etc.

Ex.: <d. o marido pela falta de pagamento da pensão> <encaminhou-se ao tribunal, pois pretendia d.>

transitivo direto

8 Rubrica: termo jurídico.

requerer judicialmente

Ex.: d. uma indenização pelo ocorrido

Fonte: Houaiss, 2007.

E, quanto ao substantivo **demandada**, entre as 9 acepções apresentadas pelo Dicionário Houaiss, a quarta recebe a rubrica **jurídico**, conforme se vê no Quadro 9.6:

Quadro 9.6 – Acepções do Dicionário Houaiss para o verbete *demandada*

4 Rubrica: termo jurídico.

processo judicial; ação, litígio, pleito

Fonte: *Ibidem*

Considerando a variedade de acepções trazidas pelo Dicionário Houaiss a essas palavras, é possível inferir que elas circulem não só nos textos jurídicos, mas em textos de linguagem comum. Entretanto, não parece ser o que ocorre com **demandada** e **demandante** nos *corpora* aqui analisados. Para essas palavras, o Dicionário traz apenas 2 acepções, uma delas com a rubrica **termo jurídico**, conforme se vê no Quadro 9.7.

Quadro 9.7 – Acepções do Dicionário Houaiss para os verbetes *demandada* e *demandante*

<i>demandado</i>	<i>demandante</i>
<p>■ adjetivo</p> <p>1 que se demandou</p> <p>■ adjetivo e substantivo masculino</p> <p>Rubrica: termo jurídico.</p> <p>2 diz-se de ou pessoa contra quem é promovida a demanda; acionado, réu</p>	<p>■ adjetivo e substantivo de dois gêneros</p> <p>1 que ou o que demanda; demandador, pleiteador</p> <p>1.1 que ou o que move uma ação judicial contra outrem ou algo</p>

Fonte: *Ibidem*

No verbete **demandante**, não consta a rubrica **termo jurídico**, mas a definição traz essa ideia, como se vê no item 1.1. De todo modo, o fato de o Dicionário trazer apenas duas acepções, sendo uma delas jurídica, evidencia que tais palavras são mais relacionadas ao âmbito jurídico do que **demanda** e **demandar**. E parece ser o que ocorre com os nossos dados aqui.

Segundo, então, a Tabela 9.28, **demandada**, em primeiro lugar, e **demandante**, em terceiro, são palavras com maior probabilidade de não serem conhecidas pelos leitores dos *corpora* aqui analisados.

Quanto ao adjetivo **homologatória** e a forma verbal **homologo**, da mesma família, vejamos, no Quadro 9.8, o que traz o Dicionário Eletrônico Houaiss no verbete **homologar**, do qual se derivam.

Quadro 9.8 – Acepções do Dicionário Houaiss para o verbete *homologar*

■ verbo

transitivo direto

1 Rubrica: termo jurídico.

tornar executório (decisão, ato etc.); decretar (autoridade judicial ou administrativa) sentença de homologação

transitivo direto

1.1 Rubrica: termo jurídico.

ratificar, confirmar juridicamente (sentença estrangeira) em território nacional

transitivo direto

2 Derivação: por extensão de sentido (*da acp. 1*).

reconhecer (algo) oficialmente

Ex.: h. um recorde, uma vitória

transitivo direto

3 Derivação: por extensão de sentido (*da acp. 1*).

aceitar, reconhecer (algo) como legítimo

Ex.: h. a sugestão dos estudantes

Fonte: Houaiss, 2007.

É também interessante observar, no verbete do Houaiss acima, que as duas primeiras definições apresentam a rubrica **termo jurídico**. E a terceira traz “reconhecer (algo) oficialmente”. Isso evidencia que o verbo **homologar** e seus derivados são palavras de teor altamente especializado. E os dados de nossas análises aqui confirmam isso. Por circularem apenas no CE, os leitores dos *corpora* aqui analisados também terão, em princípio, dificuldade de compreender o que significam **homologatória** e **homologo**.

É preciso, porém, observar novamente aqui que nossas listas são de *types* e não de lemas. Por isso, apesar de esses *types* não constarem nos *corpora* contrastados, outras formas dessas palavras ocorrem. A Tabela 9.29 traz esses dados.

Tabela 9.29 – Frequência de *Types* da Família de *demandada/demandante* e *homologar/homologatória* nos *Corpora* de Contraste e nos *Corpora* de Referência

	Família de <i>Demandada</i>	Família de <i>Homologo</i>
FSP	105	5
ZH	132	13
JM	51	3
DG	2	5
CORPOP¹¹⁶	15	5
CR1	73	1
CR2	470.341	82.810

Quanto ao *type* **intimadas**, particípio do verbo **intimar**, em suas 397 ocorrências no CE, aparece no final da sentença, como se vê no exemplo a seguir:

(9.8) *As partes consideram-se intimadas a partir da publicação da decisão, caso tenha ocorrido no prazo assinado; do contrário, a intimação terá de ser formal.*

A forma **intimadas** não consta nos demais *corpora*. Entretanto, outras formas derivadas do verbo **intimar** aparecem nos *corpora* aqui analisados, tais como: **intimados**, **intimado**, **intimou** etc. Diante disso, não consideramos que tal palavra seja realmente exclusiva do CE.

E o adjetivo **recursais**, nas 750 vezes em que ocorre, acompanha o substantivo **turmas**, na expressão *Turmas Recursais*. Como já referimos na seção 5.2.1, este nome designa o órgão judiciário que julga os recursos dos JECs. Por tal razão, cremos que a análise do adjetivo **recursais** não pode se dar isoladamente, pois sua compreensão está condicionada ao substantivo **turmas**, que ele acompanha.

¹¹⁶ O CorPop não se encontra no Sketch Engine, por isso não contabilizamos a frequência, mas apenas a ocorrência.

Como veremos a partir de agora, a expressão **turmas recursais** consta entre as multipalavras relacionadas pelo Sketch Engine nas comparações entre o CE e o CR1 e CR2.

9.4.3 CE X CR1 – MULTIPALAVRAS

Na análise das listas de palavras únicas realizadas até aqui, ocorreram casos em que mencionamos que tais palavras apareciam recorrentemente acompanhadas de outras, formando uma expressão ou sintagma terminológico. Foi o caso, por exemplo, de **inversão**, em **inversão do ônus da prova**; e **recursais**, em **turmas recursais**.

A ferramenta Sketch Engine, quando processa a lista de palavras-chave, também processa a lista de multipalavras, que são duas ou mais palavras que aparecem sempre juntas e, por isso, ganham destaque nessa comparação. Muitas são potenciais candidatos a termos.

A Tabela 9.30 traz a relação das 20 primeiras multipalavras encontradas na comparação do CE com o CR1.

Tabela 9.30 – Vinte Primeiras Multipalavras-Chave do CE x CR1

Item	Frequência (CE)	Frequência (CR1)	Frequência Relativa (CE)	Frequência Relativa (CR1)	Pontuação (Chavicidade)
danos morais	1299	0	1998,07739	0	1999,08
parte autora	960	0	1476,63916	0	1477,64
dano moral	957	0	1472,02466	0	1473,03
turmas recursais	734	0	1129,01367	0	1130,01
energia elétrica	571	0	878,29266	0	879,293
turma recursal	501	0	770,62103	0	771,621
turma recursal cível	431	0	662,94946	0	663,949
proposta de decisão	422	0	649,10596	0	650,106
sentença homologatória	421	0	647,56775	0	648,568
recurso cível	410	0	630,64795	0	631,648

juiz leigo	388	0	596,80829	0	597,808
recurso inominado	367	0	564,50684	0	565,507
danos materiais	299	0	459,91156	0	460,912
plano de saúde	283	0	435,3009	0	436,301
juizado especial	258	0	396,84677	0	397,847
caso concreto	254	0	390,69409	0	391,694
tribunal de justiça	248	0	381,46509	0	382,465
honorários advocáticos	236	0	363,00711	0	364,007
parte ré	229	0	352,23996	0	353,24
processo civil	223	0	343,01096	0	344,011

Fonte: Sketch Engine

A estrutura da Tabela 9.30 segue o padrão das Tabelas 9.24 e 9.26, relativas às palavras-chaves de um *type* só. De imediato, podemos observar a presença de **turmas recursais**, em quarto lugar em termos de frequência.

Essas expressões espelham o universo das sentenças aqui analisadas. **Danos morais**, em primeiro lugar, representa temática que ocupa grande parte dos processos dos JECs. **Parte autora**, em segundo, refere-se a uma das partes do processo; **turmas recursais** já comentamos, e assim sucessivamente.

Em princípio, quase todas podem ser consideradas termos, porque indicam um instituto jurídico relativo ao âmbito dos JECs. É o caso, por exemplo, de **danos morais** e **sentença homologatória**. Esta última significa a decisão que, grosso modo, atribui efeito jurídico ao parecer do juiz leigo.

Há outras expressões, porém, que, apesar de constarem na lista em razão da sua alta frequência, não podem ser, em tese, consideradas termos. É o caso, por exemplo, de **caso concreto** e **energia elétrica**. A expressão **caso concreto** é muito recorrente nas Sentenças para fazer referência à situação examinada naquele momento, e **energia elétrica** é um dos temas debatidos nos conflitos levados aos JECs.

Cumpramos também observar que nenhuma dessas expressões consta no CR1, que é composto apenas por textos literários.

Vejamos como ficaram as listas em relação ao CR2.

9.4.4 CE X CR2 – MULTIPALAVRAS

A Tabela 9.31 traz a lista de multipalavras relativas à comparação entre o CE e o CR2.

Tabela 9.31– Vinte Primeiras Multipalavras-Chave do CE x CR2

Item	Frequência (CE)	Frequência (CR2)	Frequência Relativa (CE)	Frequência Relativa (CR2)	Pontuação (Chavicidade)
parte autora	960	2234	1476,639	0,48326	996,209
turmas recursais	734	1243	1129,014	0,26889	890,555
turma recursal cível	431	246	662,9495	0,05322	630,403
recurso cível	410	85	630,648	0,01839	620,243
sentença homologatória	421	334	647,5678	0,07225	604,865
proposta de decisão	422	402	649,106	0,08696	598,095
juiz leigo	388	460	596,8083	0,09951	543,705
recurso inominado	367	233	564,5068	0,0504	538,371
turma recursal	501	2476	770,621	0,53561	502,484
danos morais	1299	24566	1998,077	5,31415	316,603
dano moral	957	17545	1472,025	3,79536	307,177
parte ré	229	764	352,24	0,16527	303,14
quantum indenizatório	204	474	313,7858	0,10254	285,511
juizado especial cível	220	2970	338,3965	0,64247	206,637
inscrição indevida	132	257	203,0379	0,05559	193,292
empresa ré	139	592	213,805	0,12806	190,419
relação de consumo	202	3310	310,7095	0,71602	181,646
juiz de direito	448	12953	689,0983	2,80201	181,509
ilegitimidade passiva	143	1013	219,9577	0,21913	181,242
órgãos de proteção	151	1383	232,26303	0,29896	179,58

Fonte: Sketch Engine

A Tabela 9.31 segue a estruturação da Tabela 9.30. A primeira multipalavra a aparecer é **parte autora**, segunda colocada na Tabela 9.30. **Danos morais**, primeira no *ranking* da comparação com o CR1, passa aqui para a 10ª colocação.

As multipalavras listadas, tal como ocorreu na comparação com o CR1, em sua maioria, podem ser consideradas termos. É o caso, por exemplo, de **recurso inominado**, nome dado ao recurso que ingressa nas Turmas Recursais Cíveis, e **ilegitimidade passiva**, que significa, grosso modo, que alguém não tem condições de ser considerado réu, porque não deu causa ao que é alegado contra ele.

Outras expressões listadas, porém, tal como ocorreu na comparação com o CR1, não tem, em tese, essa condição de termo. É o caso de **inscrição indevida**.

Na Tabela 9.31, assim como na Tabela 9.26, chama a atenção que as palavras-chave e as multipalavras listadas também ocorrem no CR2. Nesse sentido, cumpre referir que este *corpus* é aproximadamente 7 mil vezes maior que o CE e que os textos integrantes deste *corpus* foram extraídos de *sites* variados da internet. Entre esses *sites*, constam alguns que trazem matérias do mundo jurídico, por exemplo: boletimjuridico.com.br e migalhas.com.br.

Feita a exposição das multipalavras mais frequentes do CE resultantes das comparações com o CR1 e com o CR2, julgamos também pertinente verificar que multipalavras que se repetem nessas listas. Para tanto, adotamos o mesmo procedimento aplicado às análises até aqui realizadas, valendo-nos do recurso de identificação de elementos duplicados do programa *Microsoft Office Excel*.

A Tabela 9.32 traz as 20 primeiras multipalavras mais frequentes do CE duplicadas nas listas de multipalavras do CR1 e CR2.

Tabela 9.32 – Vinte Primeiras Multipalavras mais Frequentes do CE Duplicadas nas Listas de Multipalavras do CR1 e do CR2

Item	Frequência (CE)	Frequência (CR1 e CR2)
danos morais	1299	24566
parte autora	960	2234
dano moral	957	17545
turmas recursais	734	1243
turma recursal	501	2476
juiz de direito	448	12953
turma recursal cível	431	246
proposta de decisão	422	402

sentença homologatória	421	334
recurso cível	410	85
juiz leigo	388	460
recurso inominado	367	233
danos materiais	299	9462
plano de saúde	283	22189
juizado especial	258	10271
caso concreto	254	17484
tribunal de justiça	248	125629
honorários advocatícios	236	6173
parte ré	229	764
processo civil	223	25671

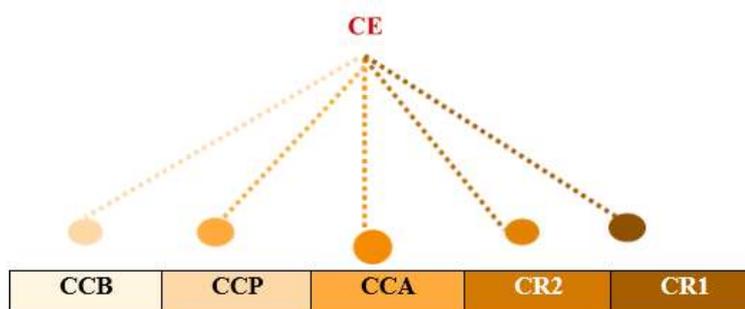
A Tabela 9.32 traz as 20 primeiras multipalavras mais frequentes do CE na comparação com o CR1 e com o CR2. Dessa lista, a única que também consta no CR1 é **juiz de direito**, com 18 ocorrências. A frequência para as demais é só relativa ao CR2. Já explicamos que isso pode se dar devido ao fato de o CR2 conter textos de *sites* jurídicos e ser muito maior que o CE, e o CR1 ser formado somente por textos literários.

De todo modo, **danos morais** (plural), em primeiro lugar, e **dano moral** (singular), em terceiro, se destacam na lista e espelham a temática tratada nos textos do CE. Conforme comentamos em relação à lista de multipalavras do CR1, é o tema que ocupa grande parte dos processos dos JECs. As demais continuam também a espelhar o universo dos JECs: **turmas recursais** (plural), **turma recursal** (singular), **sentença homologatória**, **recurso inominado**, etc.

Considerando expostos os dados relativos às comparações dos padrões lexicais do CE com os *corpora* de contraste e *corpora* de referência, ressaltamos que, por limitações de espaço e tempo, as análises aqui apresentadas se deram com os 20 primeiros itens de maior frequência nas listas geradas em cada contraste. Por serem os mais frequentes, certamente são os que melhor retratam as características principais dos padrões lexicais nos textos analisados, como vimos no decorrer do capítulo.

A seguir, então, passamos a apresentar as conclusões a que os resultados advindos das análises até aqui realizadas nos permitem chegar. Para tanto, reproduzimos a Figura 5.1, colocada no Capítulo 5 deste trabalho.

Figura 5.1 – Corpora de Pesquisa



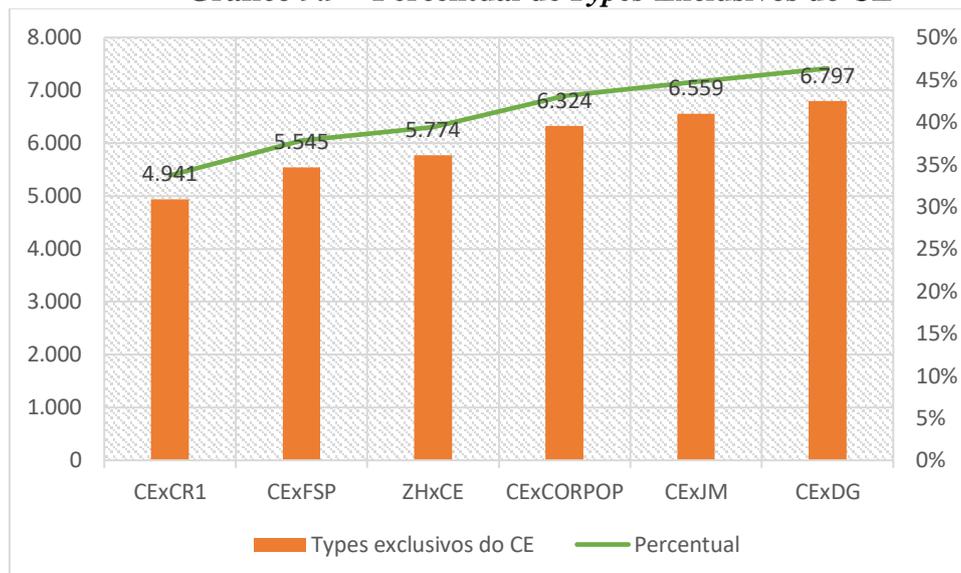
Fonte: Produção da autora

Como se vê na Figura 5.1, os *corpora* foram listados numa linha horizontal, colorida em tons gradientes. A linha começa com tons mais claros em que constam os *corpora* que representam o PPE - o CCB e o CCP. À medida que os tons vão escurecendo, aparecem os *corpora* cuja linguagem vai se aproximando de um português mais culto até chegar ao CR1, que representa o português erudito, com as obras literárias dos séculos XIX e XX.

Conforme explicamos naquele momento, nosso objetivo foi verificar para qual lado dessa linha pende o CE e, caso fosse necessário, propor alternativas para aproximá-lo de um padrão semelhante aos *corpora* representantes do PPE.

Os resultados exibidos no Gráfico 9.9¹¹⁷, reproduzido abaixo, demonstram que os padrões lexicais das Sentenças aqui analisadas estão mais próximos do CR1, com um português erudito, e mais distantes dos *corpora* que representam o PPE. Relembrando: o número colocado acima de cada coluna indica o número de *types* das Sentenças que não ocorrem nos outros *corpora*.

¹¹⁷ O CR2 não entrou nessa comparação, visto que está disponível apenas na ferramenta Sketch Engine, sem possibilidade de *download*.

Gráfico 9.9 – Percentual de *Types* Exclusivos do CE

Além disso, tais resultados confirmam algumas hipóteses de nossa pesquisa, reproduzida a seguir.

Hipótese 1

O texto da maioria das Sentenças dos JECs do PJRS apresenta uma configuração lexical, morfossintática e terminológica pouco coincidente com padrões do Português Popular Escrito (PPE). Essa configuração pode dificultar sua inteligibilidade para pessoas de escolaridade limitada e pouca experiência em leitura.

O fato de o vocabulário constante nas Sentenças aqui analisadas terem de 34% a 46% de *types* ausentes nos demais *corpora* comprova que realmente os textos da maioria das Sentenças dos JECs do PJRS apresenta uma configuração lexical e terminológica muito distante dos padrões lexicais do PPE.

Hipótese 2.b

A comparação dos textos das Sentenças dos JECs do PJRS com textos que representem o Português Popular Escrito (PPE) e também com textos de literatura dos sécs. XIX e XX, que representam o Português Escrito erudito, pode ajudar na verificação de eventuais níveis de dificuldade e/ou complexidade linguístico-textual e terminológica dessas Sentenças.

A comparação entre as listas de palavras realizadas neste capítulo permitiu encontrar palavras e expressões que, em tese, são desconhecidas do leitor dos *corpora* analisados. Tendo em conta, primeiro, que alguns *corpora* são compostos por textos de

jornais de grande circulação nos estados em que são publicados e também em nível nacional, e, segundo, que o CorPop é considerado uma referência do PPE, essas palavras têm, sim, grande chance de não serem conhecidas pela maioria da população brasileira.

Além disso, o percentual de 34% de *types* do CE ausentes no CR1 indica que inclusive o cidadão que leu as obras desse *corpus*, em princípio com maior letramento, terá grande chance de não conhecer muitas das palavras presentes nas Sentenças aqui analisadas.

Tendo conhecimento de quais são essas palavras potencialmente complexas, é possível propor alternativas de escrita com sinônimos mais simples e que se encontrem nos textos dos *corpora* analisados. Fizemos isso com **verossimilhança e hipossuficiência** e, no próximo capítulo, dedicaremos nossa atenção a este e outros processos de simplificação textual.

10 PROPOSTAS DE SIMPLIFICAÇÃO TEXTUAL

Entre os objetivos deste trabalho está também o de, com base nos dados advindos das análises dos padrões sintáticos e lexicais dos *corpora* de contraste e de referência, apresentar reescritas simplificadas para algumas das situações que se revelaram potencialmente complexas. Usamos “algumas”, porque, obviamente, não seria possível, no âmbito desta pesquisa, abarcar todas as situações.

Nossas análises contemplaram aspectos sintáticos, conforme Capítulos 7 e 8, e aspectos lexicais, conforme Capítulo 9. Os aspectos sintáticos referem-se aos constituintes frasais com função de sujeito e de complementos verbais, e os aspectos lexicais referem-se ao vocabulário apresentado em cada *corpus* analisado. As comparações realizadas nessas análises tiveram sempre como foco verificar as peculiaridades do CE em relação aos demais *corpora*, segundo os princípios da Linguística de *Corpus*.

Importa também lembrar aqui o perfil de leitor-alvo estabelecido neste trabalho: um cidadão brasileiro, adulto com escolaridade máxima de até o 9º ano do Ensino Fundamental. Esse leitor, como já mencionamos, enquadra-se entre os cidadãos brasileiros das classes C, D e E, em tese, com pouca experiência em leitura.

Na próxima seção, iniciamos a apresentação de propostas de simplificação textual.

10.1 PROPOSTAS DE REESCRITA

Das análises realizadas no Capítulo 9, chegamos à Tabela 9.28, que lista os *types* que constam apenas nas listas do CE ali apresentadas: **demandada, recursais, demandante, homologatória, homologo, intimadas**. Como trabalhamos com *types* e não com lemas, verificamos que outras formas da família dessas palavras ocorrem nos *corpora* contrastados.

Considerando que, nas tabelas apresentadas no Capítulo 9, constam apenas os 20 *types* mais frequentes em cada lista, analisamos as listas totais de *types* de todos os *corpora* de contraste, do CR1 e do CE em conjunto para encontrar outros *types* que se destacam no CE e que tenham pouca ou nenhuma frequência nos textos dos jornais analisados e no CR1. Dessa análise, chegamos aos seguintes itens:

- a) **Improcedência**: além das 425 ocorrências no CE, ocorre 1 vez no jornal ZH e 1 vez no CR1.

- b) **Quantum**: além das 372 ocorrências no CE, ocorre 2 vezes no CR1 e 6009 nos textos do Brasil constantes no CR2. Faz parte da lista de palavras-chave do CE em relação ao CR2.
- c) **In re ipsa**: além das 216 ocorrências no CE, ocorre apenas no CR2, com 222 ocorrências.
- d) **Consoante**: além das 181 ocorrências no CE, ocorre 29 vezes no CR1 e 54.987 no CR2.

Pelos dados dos itens acima, é possível observar que os *types* selecionados, com exceção de **improcedência** (1 ocorrência no jornal ZH), não constam nos *corpora* de contraste aqui analisados. Constam nos *corpora* de referência, que, por serem bem maiores que os *corpora* de contraste, têm mais probabilidade de apresentar tais itens.

A partir daqui, em novas seções, apresentamos sugestões de reescritas simplificadas para alguns desses *types* em trechos das Sentenças do CE onde se encontrem. Apresentaremos também sugestão de reescrita simplificada de uma Sentença dos JECs com apontamentos baseados em todas as análises que realizamos neste trabalho.

Nosso exercício de reformulação baseia-se em consulta prévia a fontes de apoio (dicionários e obras afins), o que pode ser um recurso para subsidiar um método ou bases de um protótipo de guia de reescrita para esse tipo textual voltado para auxiliar redatores interessados, que não tenham formação Direito.

10.1.1 Improcedência

A palavra **improcedência** em sua estrutura apresenta um prefixo de negação **in-**, que, seguido de *b* ou de *p* (procedência), transforma-se em **im-**. Isso, por si só, já é um fator complicador, porque toda negação, como vimos no Capítulo 8, requer mais da memória de trabalho do leitor. Conforme Pinker (2018), primeiro processamos a afirmação, depois a negação.

O Quadro 10.1 apresenta as acepções do dicionário Houaiss¹¹⁸ para a palavra que encabeça este verbete:

¹¹⁸ Além dos *corpora* utilizados como contraste e referência, utilizamos o Dicionário Houaiss como fonte de apoio. Ressalvamos, porém, que há muitas outras fontes disponíveis. Uma delas é o **Corpus do Português: Web/Dialects**, de Mark Davies. Disponível gratuitamente em <https://www.corpusdportugues.org/web-dial/>. Acesso em 13 jun. 2022.

**Quadro 10.1 – Acepções do Dicionário Houaiss para os verbetes
improcedência e *improcedente***

<p>Improcedência</p> <ul style="list-style-type: none"> ■ substantivo feminino 1 qualidade ou condição do que é improcedente 2 Rubrica: direito processual. falta de justificação jurídica para determinado pedido, feito à autoridade judicial, a fim de ser atendido¹¹⁹ <p>Ex.: i. de denúncia ou queixa</p>
<p>Improcedente</p> <ul style="list-style-type: none"> ■ adjetivo de dois gêneros 1 que não procede, sem fundamento Ex.: denúncia i. 2 que não forma um todo racional, lógico, conseqüente; incoerente, ilógico Ex.: argumentos i.

A acepção 1 de **improcedência** remete ao adjetivo **improcedente**, cujas acepções do Houaiss também relacionamos no Quadro 10.1. Pela acepção 2 de **improcedência**, é possível verificar o caráter terminológico desta palavra, que carrega consigo o significado que consta na acepção 1 de **improcedente** no âmbito do Direito Processual.

É possível dizer, então, que a palavra **improcedência** se confirma, pela ótica lexicográfica, como um termo no âmbito do Direito. A seguir, passamos ao exame de um exemplo de trecho em que este termo foi empregado no CE.

Trecho 1

2 – *FUNDAMENTAÇÃO:*

*É o caso de **improcedência** da demanda.*

O **trecho 1** representa o início da fundamentação de uma das Sentenças do CE. Antes há um pequeno relatório¹²⁰ do que está sendo tratado neste processo: quem pede, o que pede e contra quem. Após o título, indicando que vai começar a fundamentar a sua decisão, o juiz leigo manifesta, de imediato, no corpo do texto, a sua posição sobre o resultado final. E, para tanto, usa o termo **improcedência**. A forma como as informações

¹¹⁹ A improcedência não se dá apenas por falta de fundamentação legal, mas também por falta de prova do que se alega.

¹²⁰ Segundo o art. 38 da Lei n. 9.099/95, o relatório não é obrigatório.

são colocadas no texto, sem nenhum comentário ou explicação, permite inferir que as palavras foram empregadas com a pressuposição de que sejam de conhecimento do leitor.

Além de **improcedência**, sublinhamos o substantivo **demanda**, pois, no âmbito do CE, adquire valor terminológico significando **processo**. Tal significado é listado entre as acepções que o Dicionário Houaiss traz para essa palavra: “**4 Rubrica: termo jurídico.** processo judicial; ação, litígio, pleito”.

Esta palavra é da família de **demandada**, *type* exclusivo do CE, conforme Tabela 9.28. A forma **demanda**, porém, e outras da mesma família ocorrem nos *corpora* contrastados, de acordo com o apresentado na Tabela 9.29. Porém, é preciso observar que o *type* **demanda** pode representar a forma verbal do verbo **demandar** na 3ª pessoa do singular e também o substantivo derivado desse mesmo verbo.

Assim, podemos pressupor que **improcedência e demanda** podem oferecer, de modo recorrente, algum tipo de dificuldade de compreensão por parte dos leitores dos *corpora* aqui analisados, entre eles o nosso leitor-alvo.

Desse modo, para ampliar as possibilidades de compreensão dessa frase por esses leitores, oferecemos algumas sugestões de reescrita, constantes do Quadro 10.2. Após cada sugestão, apresentamos o escore do Índice Flesch (IF) para cada trecho – embora se precise salientar que esse índice opera em melhores condições estatísticas a partir de segmentos com 100 palavras:

Quadro 10.2 – Reescritas Sugeridas para o Trecho 1

Trecho Original	Reescrita 1
<p>É o caso (1) de improcedência (2) da demanda (3).</p> <p style="text-align: right;">IF = 60,44</p>	<p>É o caso de (1) improcedência (o pedido não tem previsão na lei ou o autor não provou o que pede) (2) da demanda (pedido) (3).</p> <p style="text-align: right;">IF = 78.35</p>
	<p>Reescrita 2</p> <p>O pedido (3) não pode ser atendido (2).</p> <p style="text-align: right;">IF = 73.54</p>
	<p>Reescrita 3</p> <p>Quem pediu perdeu.</p> <p style="text-align: right;">IF = 76,59</p>

Na Reescrita 1, após **improcedência e demanda**, inserimos entre parênteses uma explicação sobre o significado dessas palavras.

Na Reescrita 2, fizemos as seguintes alterações:

- a) Retiramos a expressão **é o caso**.
- b) Alteramos a organização sintática da frase.
- c) Substituímos a palavra **improcedência**, por **não pode ser atendido**.
- d) Substituímos a palavra **demanda** por **pedido**.

Na Reescrita 3, alteramos totalmente o texto.

Tais alterações podem ser chamadas também de estratégias simplificadoras.

Na Reescrita 1, as explicações entre parênteses após **improcedência** e **demanda** são consideradas *estratégias de simplificação lexical por explicação*. Estamos aqui trabalhando com **termos**, e nem sempre uma outra palavra usada no seu lugar vai contemplar com fidelidade tudo o que eles significam. A explicação nesses casos, além de facilitar a compreensão, poderá também apresentar ao leitor o significado do termo em questão.

Após **demanda**, colocamos **pedido**, em vez de **processo**. Assim fizemos, porque, em síntese, a ideia é dizer que *o pedido no processo não pode ser atendido*. E a palavra **pedido**, além de ocorrer com mais frequência nos *corpora* analisados, também tem seu significado contido no significado de *processo, demanda, litígio, etc.*

Na Reescrita 2, a alteração do item **a** refere-se à retirada da expressão **É o caso de**. É uma *estratégia de simplificação textual por conteúdo*.

Segundo Paraguassu (2018), a estratégia de simplificação textual por conteúdo representa, entre as estratégias simplificatórias de um texto, a mais complexa, porque é preciso verificar o que realmente importa para o leitor do texto que está sendo simplificado. Para tal procedimento, é necessário distinguir a informação principal de um texto das secundárias. No trecho em questão, a retirada do trecho parece não trazer prejuízo.

A alteração da letra **b** foi uma *estratégia de simplificação textual de ordem sintática*. Reescrevemos a frase na ordem direta: sujeito, verbo e objeto. Como já comentamos várias vezes no decorrer deste trabalho, a ordem direta das frases é um facilitador para a compreensão leitora, porque segue a ordem natural das frases na língua portuguesa.

As alterações das letras **c** e **d** são estratégias de simplificação lexical e também sintáticas, uma vez que o substantivo **improcedência** foi substituído pela oração **não pode ser atendido**.

Na Reescrita 3, utilizamos uma sumarização, em uma operação que pode ser considerada bastante radical. Afinal, temos um resumo como estratégia de simplificação textual.

Pelos escores do IF anotados abaixo de cada reescrita, é possível verificar que as simplificações apresentadas facilitaram o texto. A mais fácil foi a Reescrita 1, apesar de conter mais palavras. Depois vem a Reescrita 3 e, por fim, a Reescrita 2.

10.1.2 *Quantum*

Quantum, em latim, significa **quanto**. Nas Sentenças aqui analisadas é empregado com o significado de **quantia**. Como já anotamos, além de ocorrer no CE, este estrangeirismo também ocorre no CR1 (2 vezes) e no CR2 (6009). Mesmo assim, ficou na 16ª colocação na lista de palavras-chave do CE em relação ao CR2. Tais dados revelam a especificidade de *quantum* nos textos das Sentenças aqui analisadas.

Em 204 das suas 372 ocorrências, vem acompanhado do adjetivo indenizatório – ***quantum indenizatório***, formando um sintagma híbrido: um adjetivo em língua portuguesa – **indenizatório** - qualifica um substantivo em latim - ***quantum***. Por essa construção ser mais frequente no CE, escolhemos um trecho das sentenças em que ela ocorre para apresentarmos sugestões de reescritas simplificadas.

Trecho 2

Ultrapassada a questão acerca da configuração dos danos morais, resta se aquilatar acerca da adequação do ***quantum indenizatório*** a ser fixado.

No Trecho 2, além da expressão em foco – ***quantum indenizatório*** -, sublinhamos algumas palavras que ocorrem com muito baixa frequência, ou não ocorrem, nos *corpora* de contraste e de referência. A Tabela 10.1 traz esses dados.

Tabela 10.1 – Frequência dos *Types* sublinhados no Trecho 2

<i>CORPUS</i> <i>TYPES</i>	CE	FSP	ZH	DG	JM	CORPOP	CR1	CR2
<u>acerca</u>	157	20	5	2	5	6	88	115.331
<u>adequação</u>	27	2	2	2	3	0	0	67.056
<u>aquilatar</u>	2	0	0	0	0	0	13	1.295
<u>configuração</u>	52	2	1	1	1	1	1	115.272
<u>ultrapassada</u>	86	26	16	15	27	20	10	264.115
danos morais	Multipalavra mais frequente entre as multipalavras resultantes das comparações entre o CR1 e o CR2 (Tabela 9.31)							

Os números das células da Tabela 10.1 indicam as ocorrências em cada *type* no *corpus* relacionado. O CR2 apresenta maior número de ocorrências em todos eles, o que é esperado, uma vez que é o *corpus* com maior número de palavras. De todo modo, chama a atenção a baixa frequência dessas palavras, ou nenhuma, nos *corpora* de contraste e no CR1.

Das 21 palavras do Trecho 2, 10 foram destacadas, ou seja, quase 50% do trecho é composto de palavras que podem ser consideradas potencialmente complexas e merecedoras de atenção nas sugestões de reescritas que passamos a apresentar.

Trecho Original	Reescrita 1
<p><u>Ultrapassada</u> (1) a questão <u>acerca</u> (2) da <u>configuração</u> (3) dos <u>danos morais</u> (4), <u>resta</u> (5) se <u>aquilatar</u> (6) <u>acerca</u> (7) da <u>adequação</u> (8) do quantum indenizatório (9) a ser fixado.</p> <p style="text-align: right;">IF = 14,00</p>	<p><u>Superada</u> (1) a questão <u>sobre</u> (2) a <u>ocorrência</u> (3) dos <u>danos morais</u> (perdas por causa de ofensas ao nome, à imagem, entre outras coisas, de uma pessoa) (4), <u>é preciso avaliar como fixar</u> (5, 6, 7 e 8) a quantia devida (9).</p> <p style="text-align: right;">IF = 39,22</p>

Na Reescrita 1, utilizamos as seguintes estratégias simplificadoras:

- Em (1), substituímos **ultrapassada** por **superada**, de maior ocorrência (194 ocorrências contra 84 de **ultrapassar** no total dos jornais do CCA e CCB) nos *corpora* de contraste: estratégia de simplificação lexical por substituição.
- Em (2), substituímos **acerca** por **sobre** (3074 só nos jornais do CCA e CCB), de maior ocorrência nos *corpora* de contraste: estratégia de simplificação lexical por substituição. Tal substituição exigiu também uma alteração de ordem sintática, visto que **acerca** exige a preposição **de**, e a **sobre** não.
- Em (3), substituímos **configuração** por **ocorrência** (183 ocorrências em todos os jornais do CCA e CCB): estratégia de simplificação lexical por substituição.

Ocorrência não é um sinônimo de **configuração**, mas pode ser empregada no trecho, porque a ideia é dizer que se confirmou a ocorrência de um dano.

- d) Em (4), inserimos entre parênteses uma explicação para a expressão terminológica **danos morais**: estratégia de simplificação lexical por explicação.
- e) Em (5, 6, 7 e 8), substituímos o trecho **resta se aquilatar acerca da adequação** por **é preciso avaliar como fixar**. Aqui foram utilizadas estratégias de substituição lexical por substituição – **resta** por **é preciso** (2101 ocorrências do verbo precisar nos jornais do CCA e CCB contra 15 do verbo restar); **se aquilatar** por **avaliar** (apenas pelas 2 ocorrências do verbo **aquilatar** no CE e por sua aparição nos *corpora* de referência, percebe-se a alta probabilidade de essa palavra não ser conhecida pelo nosso leitor-alvo) -; e estratégias de simplificação lexical e sintática ao mesmo tempo - **acerca da adequação** por **como fixar**.
- f) Em (9), substituímos **quantum indenizatório** por **quantia devida**: estratégia de simplificação lexical por substituição.

Trecho Original	Reescrita 2
<p><u>Ultrapassada</u> (1) a questão <u>acerca</u> (2) <u>da configuração</u> (3) dos <u>danos morais</u> (4), <u>resta</u> (5) se <u>aquilatar</u> (6) <u>acerca</u> (7) da <u>adequação</u> (8) do quantum indenizatório (9) a ser fixado.</p> <p style="text-align: right;">IF = 14,00</p>	<p><u>Ficando comprovados os</u> (1) <u>danos morais</u> (<u>perdas por causa de ofensas ao nome, à imagem, entre outras coisas, de uma pessoa</u>) (2), <u>é preciso estabelecer</u> (3) a quantia devida (4).</p> <p style="text-align: right;">IF = 40,22</p>

Na Reescrita 2, utilizamos as seguintes estratégias simplificadoras:

- a) Em (1), substituímos **ultrapassada a questão acerca da configuração dos** por **ficando comprovados os**: estratégia de simplificação lexical por substituição e sintática ao mesmo tempo.
- b) Em (2), inserimos entre parênteses uma explicação para a expressão terminológica **danos morais**: estratégia de simplificação lexical por explicação.
- g) Em (3), substituímos **resta se aquilatar acerca da adequação do** por **é preciso estabelecer a**. Aqui foram utilizadas estratégias de substituição lexical por substituição e sintáticas ao mesmo tempo.

Trecho Original	Reescrita 3
<p>Ultrapassada (1) a questão acerca (2) da configuração (3) dos danos morais (4), resta (5) se aquilatar (6) acerca (7) da adequação (8) do <i>quantum indenizatório</i> (9) a ser fixado.</p> <p style="text-align: right;">IF = 14,00</p>	<p>Comprovados os (1) danos morais (perdas por causa de ofensas ao nome, à imagem, entre outras coisas, de uma pessoa) (2), a quantia devida (4) precisa ser fixada (3 e, 5).</p> <p style="text-align: right;">IF = 50,87</p>

Na Reescrita 3, utilizamos as seguintes estratégias simplificadoras.

- a) Em (1), substituímos **ultrapassada a questão acerca da configuração dos** por **comprovados os**: estratégia de simplificação lexical por substituição e sintática ao mesmo tempo. Entre a reescrita 2 e 3, retiramos o gerúndio **ficando**.
- b) Em (2), inserimos entre parênteses uma explicação para a expressão terminológica **danos morais**: estratégia de simplificação lexical por explicação.
- c) Em (3, 4 e 5), substituímos **resta se aquilatar acerca da adequação do** por **a quantia devida precisa ser fixada**. Aqui foram utilizadas estratégias de substituição lexical por substituição e sintáticas ao mesmo tempo.

Segundo o IF anotado após cada reescrita, é possível verificar que do trecho original até a reescrita 3, diminuiu a complexidade.

Até aqui apresentamos algumas estratégias simplificadoras de pequenos trechos. A partir de agora, apresentamos uma sentença em sua totalidade. Afinal, é “no todo do texto que se pode melhor explicar, funcional e comunicativamente, o uso linguístico especializado, a preferência por determinados recursos linguísticos”. (HOFFMANN, 2015, p. 47). Isso, de modo nenhum, invalida as sugestões simplificadoras que apresentamos. É preciso considerar que os trechos simplificados foram também extraídos das Sentenças e apresentam, como vimos, palavras que podem, de fato, oferecer alguma dificuldade de compreensão por parte não só do nosso leitor-alvo, mas também dos leitores dos textos dos *corpora* aqui analisados.

10.1.3 Exemplar de Sentença

A seguir, apresentamos um exemplo de uma Sentença completa dos JECs. Por questão de espaço, extraímos os dados do cabeçalho e alguns parágrafos. Os segmentos desconsiderados, assinalados com [...], juntamente com o cabeçalho, representam um universo de 442 palavras em relação ao texto completo, que tem 1142 palavras.

PROPOSTA DE SENTENÇA

Comarca XXXXXXXX

Juizado Especial Cível

Vistos, etc.

Em que pese a dispensa do relatório (1) consoante art. 38, *in fine*, da Lei nº 9.099/95, faz-se um breve arrazoado dos fatos (1).

Trata-se ação indenizatória promovida por xxxxxxxx em face de xxxxxxxx, em que narrou (2) ter ingressado em um consórcio, cujo o (sic) bem ofertado é um automóvel, em cem parcelas de R\$ 397,61 (trezentos e noventa reais e sessenta e um centavos) acrescidos de tarifas mensais.

Referiu (2), [...]. Aduz (2) ter entrado em contato com a requerida para esclarecer tal fato e solicitar a retirada dessa taxa de seus boletos futuros, porém não obteve sucesso. Requeveu (2) a condenação da ré a (sic) restituição em dobro do valor pago e a (sic) indenização a título de danos morais.

No mérito, a análise do cotejo probatório carreado aos autos conduz à improcedência por parte autora do pedido articulado na peça preambular.

Fundamento.

Inicialmente, calha salientar que a relação que envolve os litigantes é eminentemente de consumo, assumindo cada qual a posição de consumidor como destinatário final e prestador de serviços. Corolário lógico é a aplicação das regras previstas no CDC, inclusive a inversão do ônus probandi (1).

Consequentemente, no presente caso cabe inverter-se o ônus da prova conforme previsto no art. 6º, VIII do CDC, recaindo assim o ônus da comprovação (1) sobre a demandada.

Da análise das teses levantadas pelas partes e documentos juntados, verifica-se que a requerida agiu de boa-fé contratual (1), exibindo um compromisso de fidelidade e cooperação nas relações contratuais. Também é visível que a requerida cumpriu com o dever da informação (1), conforme documento em anexo (fls. 48).

Na opinião de Rizzato Nunes, entendemos o princípio da informação preceituado no Código de Proteção e Defesa do Consumidor como [...].

No caso em apreço, é possível verificar no preenchimento do contrato que não haveria a taxa de administração mensal a partir da parcela 50 até a 100, o que torna o primeiro erro escusável mediante raciocínio lógico.

Com relação à condenação por danos morais, tem-se que este não comporta acolhimento. Com efeito, no caso sub judice, tem-se que não restaram configurados na espécie. Em que pese não se reconheça qualquer elemento que configure situação de abalo moral e psicológico a ensejar a indenização pretendida, já que não restou configurado o descumprimento contratual.

[...].

[...].

Assim, feitas essas breves considerações, entende-se não merecer prosperar o pedido de condenação em danos morais requerido pelo autor, tendo em conta se tratar de mero dissabor a situação experimentada.

ANTE O EXPOSTO, com base na fundamentação retro e fulcro no art. 269, I, do CPC, opina-se pela IMPROCEDÊNCIA do pedido articulado pela autora, no sentido de julgar improcedente o pedido de reparação a título de danos morais. Sem sucumbência ou condenação de honorários.

Remeta-se à apreciação da Presidência. Intimem-se.

SENTENÇA HOMOLOGATÓRIA

Nos termos do art. 40 da Lei nº 9.099/95, homologo a proposta de decisão, para que produza efeitos como sentença. Sem custas e honorários, na forma da Lei. As partes consideram-se intimadas a partir da publicação da decisão, caso tenha ocorrido no prazo assinado; do contrário, a intimação terá de ser formal.

XXXXXXX, 16 de dezembro de 2015.

Juiz de Direito.

Na Sentença acima, fizemos algumas marcações:

- a) Textos sublinhados e seguidos de (1): são exemplos de sujeitos pospostos aos verbos. Conforme vimos neste trabalho, o sujeito pós-verbal oferece mais trabalho ao leitor, visto que não segue a ordem natural da língua portuguesa: sujeito-verbo-objeto.
- b) Alguns verbos na fonte azul: representam verbos com sujeitos elípticos, que foram de maior ocorrência entre os tipos de sujeitos encontrados no CE.
- c) Alguns verbos com destaque azul: são verbos que têm como referente o redator do texto, porém alguns estão flexionados na 3ª pessoa do singular e um na 1ª pessoa do plural.
- d) Alguns substantivos com fonte bordô: representam formas diferentes de fazer referência à parte contra quem se pede algo no caso. Percebe-se um esforço de não repetir a mesma palavra várias vezes.
- e) A sigla CDC em destaque verde: chama a atenção que, inicialmente, só a sigla aparece. Na terceira referência, coloca-se o nome por extenso do código que ela abrevia – Código de Defesa do Consumidor. No Capítulo 9 comentamos sobre a falta do nome por extenso nas siglas empregadas no CE.

Há, ainda, vários trechos em destaque amarelo. Nesses trechos constam palavras ou grupos de palavras que não parecem muito comuns, conforme as ocorrências nos *corpora* analisados neste trabalho postas após nossa apreciação sobre cada uma delas.

- 1) **Em que pese:** Sinônimo de **apesar, ainda que**; 85 ocorrências no CE; 11 no CR1, 1 nos jornais FSP, ZH e DG; e nenhuma no JM e no CorPop.
- 2) **Consoante:** Sinônimo de **de acordo com, conforme**. É uma das expressões listadas no início deste capítulo, por serem frequentes no CE, mas com pouca ou nenhuma ocorrência nos demais *corpora*. Mais precisamente, não ocorre nos jornais do CCA e CCB nem no CorPop, apenas 29 vezes no CR1.
- 3) **In fine:** expressão em latim que significa **no fim**.
- 4) **Arrazoado:** Significa, entre outras coisas, **exposição de razões**. Ocorre apenas 1 vez no CE e no CR1, e nenhuma nos jornais do CCA e CCB e no CorPop.
- 5) **Em face de:** expressão que, no âmbito do processo, tem um emprego bem específico e aqui é singelamente traduzida como sendo: a pessoa pede ao Estado

uma medida judicial contra alguém. Ou seja, neste caso, não é empregada com o significado usual de **em razão de**.

- 6) **Cotejo probatório carregado aos autos:** significa, grosso modo, a análise das provas trazidas ao processo. **Cotejo** ocorre apenas no CE. **Probatório** ocorre 225 vezes no CE; 2 vezes no CR1 e 3 no jornal FSP. **Carregado:** ocorre apenas no CE
- 7) **Improcedência:** já analisamos no Trecho 1, que foi simplificado.
- 8) **Pedido articulado na peça preambular:** significa, grosso modo, o pedido feito quando se entra com ação na Justiça. É o pedido realizado na petição inicial do processo. Chama a atenção neste trecho o uso de **articulado** e **peça preambular**. **Articular** ocorre 3 vezes no CE, nos jornais do CCA e CCB e 1 vez no CorPop. **Preambular** ocorre apenas no CE (1 ocorrência).
- 9) **Calha salientar:** Aqui chama a atenção o verbo **calhar**. Tem, entre outras acepções, a de **vir a propósito**, que parece ser o caso aqui. Ocorre 4 vezes no CE e 11 vezes no CR1. Nos demais *corpora* não ocorre.
- 10) **Litigantes:** Segundo o Dicionário Houaiss, é um termo jurídico que designa as partes de um processo litigioso, ou seja, um processo em que há conflito de interesses. Ocorre 38 vezes no CE e 1 vez no CR1.
- 11) **Eminentemente:** advérbio derivado de **eminente**, que significa algo muito acima dos demais, muito elevado. Palavras da família de **eminente** ocorreram 7 vezes no CE, 23 no CR1 e 3 no jornal FSP. Nos demais *corpora* não ocorreram. Estudos de complexidade textual indicam que advérbios terminados em -mente oferecem maior dificuldade de compreensão leitora.
- 12) **Corolário lógico:** Significa, grosso modo, uma conclusão. **Corolário** ocorre 7 vezes no CE, 3 no CR1 e 2 no jornal FSP. Não ocorre nos demais *corpora*.
- 13) **Inversão do ônus probandi e inverter-se o ônus da prova:** Já comentamos sobre esta expressão no Capítulo 9. Em uma das vezes empregadas aqui, houve a inclusão de um latinismo – **probandi**.
- 14) **Em apreço:** Significa o que está em discussão, sendo tratado. Ocorre 36 vezes no CE, 1 vez no CR1 e nenhuma nos jornais e no CorPoP.
- 15) **Erro escusável:** Significa erro que pode ser desculpado. Ainda que tal expressão possa, em algumas áreas do Direito, ser empregada como termo, como, por exemplo, no Direito Penal, **erro escusável** foi utilizado aqui para fazer referência a um mero equívoco de digitação.
- 16) **No caso sub judice:** Significa, grosso modo, no caso em exame.

- 17) **Não restaram configurados na espécie** [...] **configure**: Palavras da família do verbo **configurar** (que significa, entre outras coisas, **parecer**) ocorrem 490 vezes no CE; 13 no CR1 e no jornal ZH; 12 no jornal FSP; 8 no CorPop; 6 no JM e 3 no DG. A expressão **na espécie**, que significa **no caso**, ocorre 43 vezes no CE, 5 no CR1 e nenhuma nos outros *corpora*.
- 18) **A ensejar**: Significa **dar motivo, razão**. Ocorre apenas no CE, 43 vezes.
- 19) **Merece prosperar**: Expressão muito empregada nos textos jurídicos e significa **merecer aceitação, aprovação**. A expressão tal como está ocorre apenas no CE, 51 vezes. O verbo **prosperar** ocorre 45 vezes no CE, 10 no DG; 6 no jornal FSP, 5 no CR1 e 2 no jornal ZH. No JM não ocorre.
- 20) **Retro**: Empregado com o sentido de anteriormente. Ocorre 22 vezes no CE e 6 no DG e 1 vez no CR1. Não ocorre nos demais *corpora*.
- 21) **Fulcro no**: Empregado com o sentido de **com apoio no**. Ocorre 61 vezes no CE e 1 no jornal FSP. Não ocorre nos demais *corpora*.
- 22) **Sucumbência**: O substantivo e termo jurídico **sucumbência** é derivado do verbo **sucumbir**, que significa **perder, ser derrotado**. Muitas vezes o juiz define um valor que a parte perdedora tem que pagar à parte vencedora: seria a sucumbência. Nesta sentença, o juiz desconsiderou a sucumbência, até mesmo por não ter previsão na lei, especificamente nessas situações das sentenças dos JECs. O substantivo **sucumbência** ocorre 42 vezes no CE e 1 vez no jornal ZH. Não ocorre nos demais *corpora*.
- 23) **Honorários**: Significa o que um profissional liberal recebe por um serviço prestado. Nas sentenças, quase sempre vem acompanhada do adjetivo **advocatícios**, fazendo referência ao valor que o advogado deve receber por seus serviços no caso. Os *honorários advocatícios* (que também são parte da sucumbência) não podem ser aplicados na decisão de primeiro grau dos Juizados Especiais Cíveis, mas tão somente nas Turmas Recursais, pois a lei dos JECs não prevê essa possibilidade.
- 24) **Sentença Homologatória e homologo**: comentamos sobre essas expressões no Capítulo 9.

As expressões apontadas acima certamente não esgotam as ocorrências de situações que, em princípio, possam parecer estranhas ao destinatário desta Sentença: o cidadão leigo que procurou a Justiça para solucionar o conflito nela posto. Como se pode

ver, muitos dos itens acima não ocorrem (se ocorrem, é com baixíssima frequência) nos *corpora* aqui analisados.

Por que usar *in fine, probandi* ou *sub iudice*? **In fine**, por exemplo, é dispensável no trecho: a referência do artigo já basta no caso. Se se quiser mantê-la, por que não usar **ao final, no fim**?

Por que usar *sub iudice*, por exemplo, quando se pode usar **em julgamento, em questão, em análise**?

Por que usar **cotejo probatório carregado aos autos, pedido articulado na peça preambular, calha salientar, corolário lógico, fulcro**?

Tais expressões não são termos, que requerem mais cuidado ao se empregar possíveis sinônimos. O que vemos com essas construções parece ser uma tentativa de mostrar erudição, uma prática por muitos chamada de “juridiquês”. Porém, essa erudição cai por terra quando se cometem equívocos gramaticais como “cujo o”, no segundo parágrafo.

Diante disso, repetimos aqui as palavras da Ministra Ellen Gracie Northfleet já colocadas neste trabalho: “A decisão deve ter caráter esclarecedor e didático. Destinatário do nosso trabalho é o cidadão jurisdicionado, não as academias jurídicas, as publicações especializadas ou as instâncias superiores” (NORTHFLEET, 2006, p. 1).

Essa fala foi em seu discurso de posse, em 2006, 16 anos atrás. Bem recentemente, no dia 03 de dezembro de 2021, em artigo publicado no jornal Zero Hora, o Desembargador Ricardo Pippi Schmidt¹²¹, Presidente da Comissão de Inovação do TJRS - INOVAJUS, assim se manifesta:

As pessoas, de maneira geral, não compreendem as decisões judiciais. Falamos e escrevemos para especialistas, sem preocupação em sermos entendidos pelo destinatário final dos serviços judiciais: a população. Daí a necessidade da mudança.

A nossa pesquisa é com Sentenças dos JECs, em que a pessoa pode fazer o seu pedido sem assistência de um advogado (ações de até 20 salários mínimos), como já várias vezes também mencionamos no decorrer deste trabalho. Por tal razão, mais esclarecedora ainda deve ser a Sentença.

Infelizmente, a Sentença aqui analisada parece não cumprir este papel didático e esclarecedor. Com isso e tendo em conta o objetivo de nossa pesquisa, apresentamos uma

¹²¹ **A linguagem simples nos tribunais.** Artigo publicado na edição do dia 03 de dezembro de 2021 do jornal Zero Hora <https://ajuris.org.br/a-linguagem-simples-nos-tribunais-por-ricardo-pippi-schmidt/>. Acesso em: 04 maio 2022.

sugestão¹²² de reescrita simplificada para ela. Assim, a nossa apresentação do original e da reescrita por nós sugerida se dará por parágrafos, para melhor visualização.

- **Primeiro parágrafo**

Original	Reescrito
<i>Em que pese a dispensa do relatório consoante art. 38, in fine, da Lei nº 9.099/95, faz-se um breve arrazoado dos fatos.</i>	Mesmo que o art. 38 da Lei 9.099/95 não exija o relatório, faremos um pequeno resumo dos fatos.
<p>Neste parágrafo:</p> <ul style="list-style-type: none"> • Substituímos em que pese por mesmo que e reformulamos a construção da frase com a retirada das demais expressões em destaque amarelo. • Colocamos o verbo fazer na 1ª pessoa do plural, para padronizar essas situações no texto. Trata-se de uma opção, pois poderiam também ficar todos na 3ª ou 1ª pessoas do singular (forma que mais ocorre nas Sentenças dos JECs). <p>Estratégias de simplificação adotadas:</p> <ul style="list-style-type: none"> • Simplificação lexical por substituição e redução da informação • Simplificação sintática 	

- **Segundo parágrafo**

Original	Reescrito
<i>Trata-se ação indenizatória promovida por xxxxxxxx em face de xxxxxxxx., em que narrou ter ingressado em um consórcio, cujo o (sic) bem ofertado é um automóvel, em cem parcelas de R\$ 397,61 (trezentos e noventa reais e sessenta e um centavos) acrescidos de tarifas mensais.</i>	XXXXXXXX ingressou com uma ação de indenização contra , XXXXXXXX, dizendo que entrou num consórcio de um automóvel em cem parcelas de R\$ 397,61 (trezentos e noventa reais e sessenta e um centavos) mais tarifas mensais.
<p>Neste parágrafo:</p> <ul style="list-style-type: none"> • Alteramos a organização sintática da frase, deixando o sujeito antes do verbo como agente da ação, numa voz ativa. Estudos de complexidade textual recomendam que, sempre que possível, se empregue a voz ativa. • Substituímos em face de por contra, para ficar mais simples para o nosso leitor. Sabemos que há discussões acadêmicas sobre o emprego dessas expressões no âmbito do processo, porém, para os fins aqui propostos, julgamos que contra fica adequado. <p>Estratégias de simplificação adotadas:</p> <ul style="list-style-type: none"> • Simplificação lexical por substituição • Simplificação sintática 	

¹²² Frisamos que se trata de uma sugestão. Certamente pode haver outras escritas para o texto que fiquem inclusive mais simples que a ora apresentada.

- Terceiro parágrafo

Original	Reescrito
Referiu, [...]. Aduz ter entrado em contato a requerida para esclarecer tal fato e solicitar a retirada dessa taxa de seus boletos futuros, porém não obteve sucesso. Requereu a condenação da ré a (sic) restituição em dobro do valor pago e a (sic) indenização a título de danos morais.	Referiu, [...]. Conta que entrou em contato com a requerida para esclarecer tal fato e solicitar a retirada dessa taxa de seus boletos futuros, porém não obteve sucesso. Requereu a condenação da ré à restituição em dobro do valor pago e à indenização a título de danos morais.
<p>Neste parágrafo:</p> <ul style="list-style-type: none"> • Substituímos o verbo “aduz”, com ocorrência apenas no CE (135) e no CR1 (16). <p>Estratégia de simplificação adotada:</p> <ul style="list-style-type: none"> • Simplificação lexical por substituição 	

- Quarto parágrafo

Original	Reescrito
No mérito, a análise do cotejo probatório carreado aos autos conduz à improcedência por parte autora do pedido articulado na peça preambular .	No mérito, a análise das provas trazidas aos autos não permite que o pedido inicial seja atendido.
<p>Neste parágrafo:</p> <ul style="list-style-type: none"> • Suprimimos as expressões em destaque amarelo. <p>Estratégias de simplificação adotadas:</p> <ul style="list-style-type: none"> • Simplificação lexical por redução da informação e substituição 	

- Quinto e sexto parágrafos

Original	Reescrito
<p><i>Fundamento.</i></p> <p>Inicialmente, calha salientar que a relação que envolve os litigantes é eminentemente de consumo, assumindo cada qual a posição de consumidor como destinatário final e prestador de serviços. Corolário lógico é a aplicação das regras previstas no CDC, inclusive a inversão do ônus probandi (1). Conseqüentemente, no presente caso cabe inverter-se o ônus da prova conforme previsão do art. 6º, VIII do CDC, recaindo assim o ônus da comprovação sobre a demandada.</p>	<p>Fundamento.</p> <p>Inicialmente, destaca-se que a relação entre as partes deste processo é de consumo: uma das partes é o consumidor, e a outra é o prestador de serviços. Portanto, as regras do Código de Defesa do Consumidor (CDC), inclusive a inversão do ônus da prova, devem ser aplicadas no caso. A inversão do ônus da prova significa que quem tem que provar é a parte mais forte na relação, e não obrigatoriamente quem alega. Por isso,</p>

	seguindo o que diz o inc. VIII do art. 6º do CDC, quem deverá provar neste processo é a empresa responsável pelo consórcio.
<p>Nestes parágrafos:</p> <ul style="list-style-type: none"> • Substituímos, nas duas primeiras frases, as expressões em destaque amarelo por outras mais simples ou suprimimos, alterando a estrutura das orações. • Acrescentamos o significado da sigla CDC e uma explicação para a expressão terminológica inversão do ônus da prova. • Suprimimos o latim <i>probandi</i>. <p>Estratégias de simplificação adotadas:</p> <ul style="list-style-type: none"> • Simplificação lexical por substituição • Simplificação lexical por explicação • Simplificação sintática 	

- **Sétimo e oitavo parágrafos**

Original	Reescrito
<p><i>Da análise das teses levantadas pelas partes e documentos juntados, verifica-se que a requerida agiu de boa-fé contratual, exibindo um compromisso de fidelidade e cooperação nas relações contratuais. Também é visível que a requerida cumpriu com o dever da informação, conforme documento em anexo (fls. 48). Na opinião de Rizzato Nunes, entendemos o princípio da informação preceituado no Código de Proteção e Defesa do Consumidor como [...].</i></p>	<p>Da análise das razões trazidas pelas partes e documentos juntados, verifica-se que a empresa responsável pelo consórcio agiu de boa-fé, exibindo um compromisso de fidelidade e cooperação nas relações contratuais. Também cumpriu com o dever da informação, conforme documento em anexo (fl. 48). Na opinião de Rizzato Nunes, o princípio da informação estabelecido no CDC é entendido como [...].</p>
<p>Nestes parágrafos:</p> <ul style="list-style-type: none"> • Deixamos a frase em destaque amarelo mais direta, suprimindo algumas palavras. • Alteramos a estrutura da última frase • Deixamos apenas CDC, porque o nome por extenso já foi colocado anteriormente • Inserimos a expressão o princípio da informação [...] como sujeito de é entendido. <p>Estratégias de simplificação adotadas:</p> <ul style="list-style-type: none"> • Simplificação lexical por redução da informação • Simplificação sintática 	

- **Nono parágrafo**

Original	Reescrito
<p>No caso em apreço, é possível verificar no preenchimento do contrato que não haveria a taxa de administração mensal a partir da parcela 50 até a 100, o que torna o primeiro erro escusável mediante raciocínio lógico.</p>	<p>No caso, é possível verificar no contrato que não haveria a taxa mensal a partir da parcela 50 até a 100. Assim, o primeiro erro apontado não se deu de forma proposital, como se verifica com a leitura do texto.</p>
<p>Neste parágrafo:</p> <ul style="list-style-type: none"> • Retiramos as expressões em destaque amarelo. • Transformamos a frase em 2, com o uso do articulador assim. • Substituímos o trecho a partir de escusável por uma redação em princípio mais direta e simples e coerente. O que seria erro escusável mediante raciocínio lógico? <p>Estratégia de simplificação adotada:</p> <ul style="list-style-type: none"> • Simplificação lexical por redução da informação • Simplificação sintática 	

- **Décimo Parágrafo:**

Original	Reescrito
<p>Com relação à condenação por danos morais, tem-se que este não comporta acolhimento. Com efeito, no caso sub judice, tem-se que não restaram configurados na espécie. Em que pese não se reconheça qualquer elemento que configure situação de abalo moral e psicológico a ensejar a indenização pretendida, já que não restou configurado odescumprimento contratual.</p>	<p>Quanto ao pedido de condenação por danos morais, ele não pode ser atendido, porque esses danos não foram comprovados no processo. Não existe prova de abalo moral e psicológico para se exigir a indenização solicitada. A empresa responsável pelo consórcio cumpriu o contrato.</p>
<p>Neste parágrafo:</p> <ul style="list-style-type: none"> • Aglutinamos as duas primeiras frases numa só, com a retirada ou substituição dos trechos em destaque amarelo. A expressão em que pese estabelece uma ressalva na frase que inicia, mas, na continuidade, não se consegue encontrar o que é ressaltado por ela. • Apresentamos, a partir de Não existe (no parágrafo reescrito), uma redação que tenta melhor traduzir a ideia do trecho em questão. <p>Estratégias de simplificação adotadas:</p> <ul style="list-style-type: none"> • Simplificação lexical por redução da informação • Simplificação sintática 	

• **Décimo Primeiro parágrafo**

Original	Reescrito
<p>[...]. [...]. Assim, feitas essas breves considerações, entende-se não merecer prosperar o <u>pedido de condenação em danos morais requerido pelo autor</u>, tendo em conta se tratar mero dissabor a situação experimentada.</p>	<p>[...]. [...]. Assim, feitas essas breves considerações, entendemos que o pedido de condenação em danos morais solicitado pelo autor não pode ser atendido. A situação ocorrida aqui representou mero aborrecimento.</p>
<p>Neste parágrafo:</p> <ul style="list-style-type: none"> • Colocamos, por padronização, o verbo na 1ª pessoa do plural. • Trouxemos o sujeito o pedido de condenação em danos [...] para antes do verbo, que foi substituído por não pode ser atendido. <p>Estratégias de simplificação adotadas:</p> <ul style="list-style-type: none"> • Simplificação lexical por substituição • Simplificação sintática 	

• **Décimo Segundo Parágrafo**

Original	Reescrito
<p>ANTE O EXPOSTO, com base na fundamentação retro e fulcro no art. 269, I, do CPC, opina-se pela IMPROCEDÊNCIA do pedido articulado pela autora, no sentido de julgar improcedente o pedido de reparação a título de danos morais. Sem sucumbência ou condenação de honorários. Remeta-se à apreciação da Presidência. Intimem-se.</p>	<p>ANTE O EXPOSTO, com base na fundamentação acima e com base no art. 269, I, do CPC, negamos o pedido de reparação a título de danos morais. Sem sucumbência ou condenação de honorários. Ou seja, quem perdeu não precisa pagar a quem venceu e também não é preciso pagar nada para o advogado. Envie-se à apreciação da Presidência. Intimem-se.</p>
<p>Neste parágrafo:</p> <ul style="list-style-type: none"> • Substituímos as expressões retro e fulcro. • Colocamos o verbo na 1ª pessoa do plural, novamente por padronização. • Acrescentamos explicações para os termos improcedência, sucumbência e honorários. A lei dos JECs não prevê sucumbência nem honorários. • Substituímos remeta-se por envie-se. <p>Estratégias de simplificação adotadas:</p> <ul style="list-style-type: none"> • Simplificação lexical por substituição e por explicação • Simplificação sintática 	

Tentamos aqui fazer um exercício de simplificação de “apenas” uma Sentença na sua totalidade. Obviamente que as reescritas apresentadas aqui se prestam a novos processos de simplificação, de modo a tornar o texto cada vez mais acessível ao nosso leitor-alvo. Além disso, nas palavras de Paraguassu (2018, p. 144),

não existe uma fórmula ‘mágica’ que garanta o sucesso de uma simplificação textual. Existem, sim, apoios para a tomada de decisão do tradutor/redator. Por exemplo: testes que podem ser realizados após a simplificação, de modo a verificar se os índices anteriores, pré-simplificação, foram alterados, reduzindo-se a complexidade do texto de acordo com as métricas referidas. Contudo, conforme antes mencionado, existem muitas variáveis implicadas nos perfis de leitores e somente uma testagem empírica (redator+ texto(s) reescrito(s)+ leitor-alvo) poderá aferir de maneira mais precisa se o resultado da simplificação produziu um texto suficientemente acessível para o leitor em questão.

A seguir, apresentamos os parágrafos reescritos na configuração da sentença original também para melhor visualização do seu todo.

PROPOSTA DE SENTENÇA

Comarca XXXXXXXX

Juizado Especial Cível

Vistos , etc.

Mesmo que o art. 38 da Lei 9.099/95 não exija o relatório, faremos um pequeno resumo dos fatos.

XXXXXXXX ingressou com uma ação de indenização contra XXXXXXXX, porque entrou num consórcio de um automóvel em cem parcelas de R\$ 397,61 (trezentos e noventa reais e sessenta e um centavos) mais tarifas mensais.

Referiu, [...]. Conta que entrou em contato a requerida para esclarecer tal fato e solicitar a retirada dessa taxa de seus boletos futuros, porém não obteve sucesso. Requereu a condenação da ré à restituição em dobro do valor pago e à indenização a título de danos morais.

[...]

No mérito, a análise das provas trazidas aos autos não permite que o pedido da petição inicial seja atendido.

Fundamento.

Inicialmente, salienta-se que a relação entre as partes deste processo é de consumo: uma das partes é o consumidor, e a outra é o prestador de serviços. Portanto, as regras do Código de Defesa do Consumidor (CDC), inclusive a inversão do ônus da prova, devem ser aplicadas no caso. A inversão do ônus da prova significa que quem tem que provar é a parte mais forte na relação, e não obrigatoriamente quem alega. Por isso, seguindo o que diz o inc. VIII do art. 6º do CDC, quem deverá provar neste processo é a empresa responsável pelo consórcio.

Da análise das razões trazidas pelas partes e documentos juntados, verifica-se que a empresa responsável pelo consórcio agiu de boa-fé, exibindo um compromisso de fidelidade e cooperação nas relações contratuais. Também cumpriu com o dever da informação, conforme documento em anexo (fl. 48).

Na opinião de Rizzato Nunes, o princípio da informação estabelecido no CDC é entendido como [...].

No caso, é possível verificar no contrato que não haveria a taxa mensal a partir da parcela 50 até a 100. Assim, o primeiro erro apontado não se deu de forma

proposital, conforme se verifica da leitura do texto.

Quanto ao pedido de condenação por danos morais, ele não pode ser atendido, porque tais danos não foram comprovados no caso. Não existe qualquer elemento que represente situação de abalo moral e psicológico para se exigir a indenização solicitada. O contrato foi cumprido.

[...].

[...].

Assim, feitas essas breves considerações, entendemos que o pedido de condenação em danos morais solicitado pelo autor não pode ser atendido. A situação ocorrida aqui representou mero aborrecimento.

ANTE O EXPOSTO, com base na fundamentação acima e com base no art. 269, I, do CPC, negamos o pedido de reparação a título de danos morais.

Sem sucumbência ou condenação de honorários. Ou seja, quem perdeu não precisa pagar a quem venceu e também não é preciso pagar nada para o advogado.

Envie-se à apreciação da Presidência.

Intimem-se.

SENTENÇA HOMOLOGATÓRIA

Nos termos do art. 40 da Lei nº 9.099/95, homologo a proposta de decisão, para que produza efeitos como sentença. Sem custas e honorários, na forma da Lei. As partes consideram-se intimadas a partir da publicação da decisão, caso tenha ocorrido no prazo assinado; do contrário, a intimação terá de ser formal.

XXXXXXX, 16 de dezembro de 2015.

Juiz de Direito.

Para verificarmos se a nossa reescrita facilitou de alguma forma o texto, submetemos ambas as sentenças – a original e a reescrita apresentada – ao sistema NILC-Metrix (Coh-Metrix 3.0, de 2020). Das métricas avaliadas por essa medição, selecionamos o Índice Flesch e a Estatística de Honoré, ambos relativos à condição de leitura do texto.

Tabela 10.2 – Índices do Coh-Metrix na Sentença Original e na Reescrita

Métrica	Sentença Original	Sentença Simplificada
Índice Flesch	43,90	57,45
Honoré	14,74	10,40

Em relação ao Índice Flesch, quanto maior o seu escore, mais fácil é o texto. A sentença reescrita passou do escore de 43,90 para 57,45, o que é bastante significativo. Segundo a Tabela 4.1 deste trabalho, textos abaixo do escore 50, são considerados difíceis. Na faixa de 50 a 75 são considerados fáceis e apropriados para o Ensino Fundamental (6º ao 9º ano).

A Estatística Honoré mede a relação entre os *tokens*, os *hapax legomena* (palavras que aparecem uma única vez no texto) e as palavras lexicais, ou palavras de conteúdo, do texto. Quanto maior é o índice desta métrica, mais rico lexicalmente é o texto. No caso, a nossa reescrita apresenta um escore menor que a original. Ou seja, tem um vocabulário menos variado e, portanto, em tese, tende a ser mais simples e mais fácil também.

Obviamente que a complexidade de um texto não se mede apenas por essas métricas, há muitas outras, como apresentamos no Capítulo 8 deste trabalho. De todo modo, eles já trazem um indicativo de que houve uma facilitação na reescrita apresentada para a sentença em questão. Ainda assim, restam vários pontos a descoberto na proposta de facilitação, sobretudo quando se pensar na interpretação de sentido para um item como o adjetivo **homologatório**. O desafio envolvido não é pequeno e se precisa considerar, ainda, a validação da reescrita por um especialista do Direito¹²³.

Este exercício na totalidade de uma Sentença judicial comprova o título de nosso trabalho: *Sentenças Judiciais: um encontro possível e necessário*. É possível, sim, este encontro. E não só é possível, como é **extremamente necessário**, principalmente no âmbito dos JECs.

Entretanto, ainda que esse encontro seja possível, o processo para que ele ocorra não se dá “num passe de mágica”. Como vimos nos exercícios aqui realizados, utilizamos várias

¹²³ Em nosso caso, tivemos a orientação do marido e Bacharel em Direito Jorge Ricardo de Lima Coelho, OAB/RS n. 40.000.

estratégias: substituímos palavras, acrescentamos explicações a alguns termos, suprimimos informações, alteramos a ordem dos elementos da frase etc.

Para a substituição de algumas palavras, por exemplo, nos valem das frequências das palavras em análise nos *corpora* aqui utilizados. Ter uma noção da frequência de uma palavra num determinado conjunto de textos é um norte para o redator. Se a palavra ocorrer com frequência em textos que são, em tese, mais acessados pela população ou pelo leitor-alvo que se tem em mente, mais probabilidade tem de ser conhecida. Com isso, a compreensão do texto em que se encontra é facilitada.

Foi isso que fizemos. Além dos *corpora* compilados por nós – 3 de jornais dos anos de 2019 e 2020 e o CR1, com obras literárias dos séculos XIX e XX -, utilizamos o do DG, de 2008 e 2010, e o CorPop. Este último, o CorPop, como já mencionamos neste trabalho, é utilizado como referência do PPE. É um acervo de vocabulário simples disponível para acesso na internet¹²⁴.

E o que vimos aqui é que, muitas das palavras do CE não se encontram nos *corpora* contrastados. Aliás, o Capítulo 9 ilustra estatisticamente esse fenômeno.

Percebemos, então, que, realmente, simplificar um texto não é uma tarefa simples, ainda que seu objetivo seja atingir uma simplicidade. Por isso, na próxima seção, relacionamos algumas orientações para uma escrita simples e/ou reescrita simplificada que podem vir a subsidiar um protótipo de um guia de redação mais acessível para esse tipo de documento.

10.2 ORIENTAÇÕES BÁSICAS PARA ESCREVER EM LINGUAGEM SIMPLES

Considerando os dados resultantes das análises que realizamos neste trabalho e considerando as orientações constantes em estudos sobre a Linguagem Simples, apresentamos algumas recomendações de escrita simples pontuais. A ideia aqui é trazer uma base inicial de subsídios para o trabalho de redatores – pessoas não especialistas em Direito, mas que trabalham em conjunto com seus operadores. Essas orientações, básicas, não só se aplicam a Sentenças judiciais como as do JECs, mas também poderiam ser estendidas a textos de qualquer outra ordem. No entanto, como temos um foco e uma série de evidências destacadas ao longo desta tese, é importante situar que os exemplos e sugestões são extraídos das Sentenças que descrevemos e analisamos. Ainda assim, os seus princípios são aplicáveis a qualquer texto que precise ser bem compreendido por seu leitor.

¹²⁴ Repetimos aqui o link de acesso ao CorPop: <http://www.ufrgs.br/textecc/porlexbras/corpop/index.php>

10.2.1 Empatia – Coloque-se no lugar do leitor

A orientação acima é fundamental para qualquer texto, mais ainda se quisermos escrever em Linguagem Simples. Considerar o leitor é traçar um perfil desse leitor: qual é o grau de instrução, em que contexto vive, que experiências em leitura pode ter, etc. E traçar este perfil é adotar uma atitude de empatia, de colocar-se no lugar desse leitor e tentar escrever para esse leitor.

Neste trabalho, delimitamos o nosso perfil de leitor: um cidadão brasileiro, adulto com escolaridade máxima de até o 9º ano do Ensino Fundamental.

Importante: Caso não consiga traçar com exatidão esse perfil de leitor, escreva pensando nas **pessoas que terão mais dificuldade de ler**. Esta é uma das razões para pesquisas em *corpora* que representem um português mais simples.

10.2.2 Organize o texto

Organizar um texto é, entre outras coisas:

- *agrupar informações que se relacionam e não misturar aquelas que tratam de temas diferentes;*
- *colocar primeiro as informações mais importantes, depois as secundárias.*

As Sentenças judiciais, não só as dos JECs¹²⁵, apresentam uma macroestrutura estabelecida em lei. Conforme o art. 489, incisos I, II e III, do Novo Código de Processo Civil (BRASIL, 2002), a Sentença judicial deve ser composta pelos seguintes elementos: o relatório, a fundamentação e o dispositivo. Mesmo assim, é preciso atentar para o próximo item.

- *ordenar as ideias nos parágrafos*

Ainda que haja essa estrutura pré-determinada nas Sentenças judiciais, **é preciso cuidar a construção dos parágrafos internos** em cada uma dessas partes. Nos textos jurídicos, parágrafos muito extensos são recorrentes. Esses parágrafos longos trazem várias ideias que, apesar de se relacionarem, tornam confuso o trecho. É o caso do exemplo ii, apresentado na introdução deste trabalho e reproduzido aqui como 10.1.

(10.1) Contudo, creio que tal cláusula não serve para afastar o direito da autora na medida em que não foi redigida de forma clara, pois não há prova de que a ré tenha

¹²⁵ Nas Sentenças dos JECs, de acordo com o art. 38 da Lei 9.099/95, o relatório não é obrigatório.

esclarecido à consumidora expressamente que a cirurgia em exame estaria excluída de atendimento do plano de saúde, sendo que o capítulo que trata das exclusões do atendimento são extremamente obscuras (sic), já que não esclarecem quais seriam os procedimentos previstos na referida Tabela INAMP/SUS, não havendo prova de que a consumidora teve vista sobre ela, razão pela qual não se vincula a tal exclusão contratual.

No exemplo acima, sublinhamos algumas palavras e/ou expressões que podem ser desconhecidas do nosso leitor-alvo. Não vamos analisar essas situações aqui, apenas sinalizamos como palavras potencialmente complexas. Além do vocabulário, temos uma frase de 98 palavras, com ideias postas de modo confuso. A memória de trabalho do leitor não capta tudo isso. Então, ao chegar ao final do trecho, provavelmente o leitor terá de voltar a ler o texto, porque “perdeu o fio da meada”.

Para maior clareza **quanto à organização (não vamos mexer no vocabulário aqui)** do parágrafo, ele poderia ficar da seguinte forma, como consta no exemplo 10.2.

- (10.2) **Contudo**, creio que tal cláusula não afasta o direito da autora porque não foi redigida de forma clara. **Não há prova** de que a ré tenha esclarecido a consumidora de que o plano de saúde não cobria a cirurgia em exame. **O capítulo** que trata das exclusões do atendimento é confuso, pois não esclarece quais os procedimentos previstos na referida Tabela INAMP/SUS. **Além disso**, não há prova de que a consumidora teve conhecimento dessa tabela. **Assim**, essa tabela não pode ser vinculada à referida exclusão contratual.

Sinalizamos em negrito o início das 5 frases que este parágrafo passou a ter. Com frases mais curtas, ainda que no mesmo parágrafo, as ideias ficam mais bem organizadas e, conseqüentemente, mais claras. Salientamos que, para o leitor-alvo desta pesquisa, esta reescrita deveria ainda passar por outros procedimentos simplificadoros, como a substituição das palavras complexas por outras mais simples.

10.2.3 Evite excesso de remissões

Empregamos a palavra **remissão** aqui para nos referirmos ao recurso linguístico utilizado para se fazer referência a elementos que já foram ou que serão citados no texto. É o caso do exemplo a seguir, em que o pronome **elas** se refere **às regras do Código de Defesa do Consumidor**, posto anteriormente no trecho:

(10.1) Embora aplicáveis **as regras do Código de Defesa do Consumidor** à relação havida entre as partes, dentre **elas** a inversão do ônus da prova, imprescindível, pois, a presença da verossimilhança nas alegações.

O emprego deste recurso muitas vezes causa ambiguidade e pode oferecer dificuldade para uma pessoa com pouca experiência em leitura.

No exemplo (7.1), repetido aqui, temos uma situação de ambiguidade.

(7.1) [...], desse modo, ao assim decidir, adotou posicionamento consentâneo com a jurisprudência desta egrégia Corte, que se orienta no sentido de considerar que, em se tratando de contrato de adesão submetido às regras do CDC, a interpretação de suas cláusulas deve ser feita da maneira mais favorável ao consumidor, [...].

No trecho “desse modo, ao assim decidir, adotou posicionamento consentâneo com a jurisprudência desta egrégia Corte, **que** se orienta no sentido de considerar”, é possível encontrar três referentes para o **que**: *posicionamento*, *jurisprudência* ou *Corte*.

Na reescrita que apresentamos em 7.1b, retiramos este **que** e iniciamos nova frase com **E este posicionamento**.

(7.1b) [...]. **Adotou**, desse modo, posicionamento consentâneo com a jurisprudência desta egrégia Corte. **E este posicionamento** é o de que, em se tratando de contrato de adesão submetido às regras do CDC, a interpretação de suas cláusulas deve ser feita da maneira mais favorável ao consumidor. [...]

10.2.4 Evite que seu leitor tenha de procurar informações fora do texto.

Nas Sentenças judiciais, a citação de artigos de lei para fundamentar o posicionamento adotado é muito frequente. Em relação às Sentenças dos JECs, universo do nosso trabalho, é preciso considerar que grande parte de seus destinatários desconhece a lei e a forma como procurar esses textos. É também preciso considerar que a linguagem empregada nas leis nem sempre é simples, pois, entre outras coisas, há abundância de termos técnicos. Desse modo, o ideal seria que o trecho fosse transcrito e, se possível, que houvesse um pequeno esclarecimento sobre o que nele está escrito.

10.2.5 Use palavras simples (mais familiares e menos complexas)

Esta é uma das primeiras e principais orientações dadas para se empregar uma linguagem simples. No decorrer de nosso trabalho, exploramos esta questão importantíssima, e

no **Quadro 10.3** retomamos alguns exemplos trabalhados, com sugestão de substituição por palavras mais simples:

Quadro 10.3 – Exemplos de palavras a evitar e substituir

EVITE	EMPREGUE
aquilatar	avaliar
calha salientar	é preciso destacar/destaca-se
caso <i>sub judice</i>	caso em exame
peça preambular	petição inicial

10.2.6 Evite palavras desnecessárias.

Em relação a esta orientação, chamamos a atenção para uma expressão muito recorrente nas Sentenças dos JECs que, em princípio, é desnecessária: **Sem custas e honorários**.

Conforme o art. 55 da Lei n. 9.099/95, só incidirão custas e honorários de advogado quando houver litigância de má-fé. Portanto, julgamos que tal expressão é desnecessária quando não ocorrer casos de má-fé, que representam a maior parte dos processos desses juizados.

Nos textos jurídicos há ainda outras expressões sem nenhuma contribuição para os textos. Elas são empregadas simplesmente por serem consideradas praxe forense. É o caso de, por exemplo: **vistos, etc.; recebidos e examinados os autos, diligências legais, etc.**

10.2.7 Use mais verbos

Empregue, sempre que possível, verbos no lugar de substantivos ou adjetivos. O verbo, por normalmente expressar ação, deixa a ideia mais direta.

Além disso, nossa mente apresenta algumas particularidades:

- sempre procura um agente: **quem/o que faz o quê**.
- entende mais facilmente coisas que os olhos possam ver ou que estejam em nossa imagem mental.

Uma das métricas do sistema NILC-Metrix¹²⁶ (Coh-Metrix 3.0, de 2020). utilizadas neste trabalho foi a relativa à presença de palavras concretas nos *corpora* em contraste. Justificamos essa escolha porque a terminologia do Direito é altamente abstrata. E o emprego de palavras abstratas é um fator complicador para a compreensão, porque não se consegue ter

¹²⁶ Disponível em: <http://fw.nilc.icmc.usp.br:23380/nilcmetrix>,

uma imagem mental do que elas significam. É mais fácil ter uma imagem mental de **mesa**, **caneta** (substantivos concretos) do que **demanda**, **homologação** (substantivos abstratos), etc.

Diante disso:

EVITE	EMPREGUE
A ré apresentou contestação .	A ré contestou .
O autor trouxe comprovação .	O autor comprovou .
A finalidade da reparação do dano moral é oferecer compensação ao lesado, atenuando seu sofrimento.	Reparar o dano moral tem por fim compensar o lesado, diminuindo seu sofrimento.

Importante:

Salientamos, porém, que o uso de substantivos abstratos serve, muitas vezes, como recurso para diluir, de modo proposital, a ação do agente. E isso ocorre muito no mundo jurídico. Um advogado de defesa, por exemplo, para retirar o peso da ação sobre o seu cliente, pode nominalizar essa ação. Em vez de:

O réu **disparou** contra a vítima na saída do restaurante.

Escreve:

O **disparo** contra a vítima ocorreu na saída do restaurante.

10.2.8 Use frases mais curtas.

O emprego de frases longas é recorrente nos textos jurídicos. Essa característica, somada ao emprego de palavras complexas, aumenta a complexidade da linguagem jurídica. Comentamos sobre esta situação no item sobre a organização dos parágrafos. A seguir, apresentamos mais um exemplo das sentenças dos JECs.

- (10.2) A tese de que houve o pedido de cancelamento do cartão de crédito em 2014 por ter deixado de receber o seu salário por meio daquela conta bancária, não se sustenta, visto que não há comprovação de tal pedido e, mesmo que o tivesse havido, o autor apresentou comportamento contrário, continuando a movimentar a conta e utilizar o seu cartão, conforme atestam os extratos bancários juntados.

Esta frase é composta de 66 palavras. Sem maiores alterações, poderia ser dividida em duas, inserindo-se um ponto-final depois de pedido. É o que fazemos na reescrita 10.2a, a seguir:

(10.2a) A tese de que houve o pedido de cancelamento do cartão de crédito em 2014 por ter deixado de receber o seu salário por meio daquela conta bancária, não se sustenta, visto que não há comprovação de tal pedido. **Mesmo que** houvesse, o autor apresentou comportamento contrário, pois continuou a movimentar a conta e a utilizar o seu cartão, conforme atestam os extratos bancários juntados.

10.2.9 Coloque, sempre que possível, a frase em ordem direta (sujeito-verbo-objeto).

Comentamos várias vezes no decorrer deste trabalho sobre a ordem canônica da língua. É a ordem mais natural e mais facilmente compreendida pela falante dessa língua. Por isso, uma das recomendações para quem quer escrever em linguagem simples é colocar a frase em ordem direta.

A seguir trazemos um exemplo das Sentenças em que o sujeito está após o verbo e sua reescrita na ordem S-V-O.

(10.3) Procedem parcialmente os pedidos da autora.

(10.3a) Os pedidos da autora procedem parcialmente.

Importante:

No decorrer de nossas análises, também comentamos várias vezes sobre alguns verbos que expressam comandos ao final da sentença. É caso do seguinte exemplo, reproduzido aqui:

(9.7) Nos termos do art. 40 da Lei n. 9.099/95, **remetam-se os autos** ao Juiz de Direito Presidente deste Juizado Especial Cível para a devida homologação. **Registre-se, Publique-se e Intimem-se.**

Tais verbos estão na voz passiva pronominal, que leva o sujeito para depois do verbo. São situações bem específicas desses textos. O que registramos aqui são recomendações, não obrigações. Assim, cada situação deve ser avaliada de acordo com o contexto.

As orientações acima não esgotam as recomendações que constam em trabalhos e manuais relativos à Linguagem Simples. Colocamos aqui as que consideramos mais pertinentes aos dados coletados neste trabalho e ao tipo de discurso/texto envolvido.

Conforme mencionamos na introdução deste trabalho, muitos tribunais brasileiros já estão, desde 2021-22, adotando os princípios da Linguagem Simples em suas normativas. É o caso do Tribunal de Justiça do Estado do Rio Grande do Sul, âmbito dos JECs, cujas Sentenças

analisamos sob a forma de *corpus* de estudo. Esse tribunal, pelo ato Ato nº 024/2022-P, de 25-04-2022, instituiu a política de utilização de Linguagem Simples.

Tal Ato foi resultado do trabalho da Comissão de Inovação do TJRS - INOVAJUS, responsável pela elaboração e edição do *Guia de Linguagem Simples*¹²⁷ desse tribunal. Essa construção se deu com apoio de vários servidores sob a coordenação de juízes e desembargadores, em especial o Des. Ricardo Pippi Schmidt (Presidente) e o Dr. André Luís Aguiar Tescheiner, que nos deram a honra de também fazer da construção desse Guia.

Na produção desse material, tivemos o privilégio de prestar apoio relativo aos aspectos teóricos e práticos da LS e da ATT. E isso fizemos com aproveitamento direto de resultados desta pesquisa de doutorado. Foi, assim, uma forma de levar os estudos acadêmicos para “a vida real” desse cenário comunicativo e discursivo. E este é, por excelência, o objetivo das pesquisas científicas, ou seja, melhorar a vida da sociedade com seus achados.

A seguir, apresentamos, a título de ilustração, algumas páginas do Guia.

¹²⁷ O trabalho se insere no âmbito do Projeto Descomplica, incluído no Planejamento Estratégico do TJRS (Portaria n. 106/2020-P). Em breve, o guia estará disponível na página do TJRS

Figura 10.1 – Capa e contracapa do Guia de Linguagem Simples do TJRS

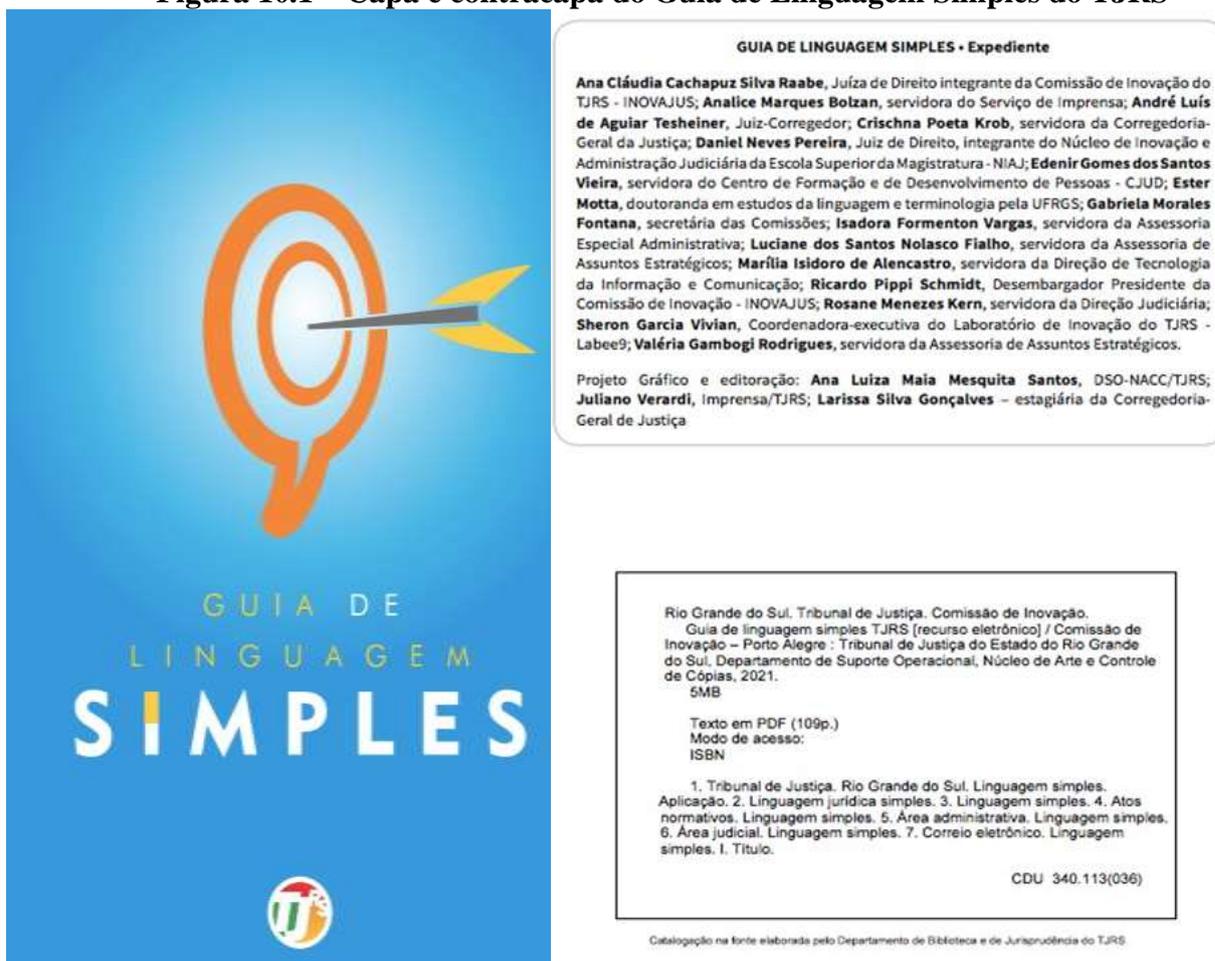


Figura 10.2 – Página 2 do Guia de Linguagem Simples do TJRS

Guia de Linguagem Simples TJRS

A propósito, o Manual de Redação da Presidência da República (BRASIL, 2018, p. 21) traz as seguintes recomendações:

- a língua culta é contra a pobreza de expressão, e não contra a sua simplicidade;
- o uso do padrão culto não significa empregar a língua de modo rebuscado ou utilizar figuras de linguagem próprias do estilo literário.

Essas orientações vão no mesmo sentido de outras dispostas nos manuais sobre a Linguagem Simples:*

A Linguagem Simples permite ao leitor

- Encontrar facilmente a informação que procura
- Compreender a informação que encontrou
- Usar a informação encontrada

Em inglês, Linguagem Simples é chamada de *Plain Language*. O movimento denominado *Plain Language* surgiu lá pela metade do século XX nos Estados Unidos e depois foi se expandindo para outros países. Para saber mais sobre a história da Linguagem Simples, acesse [este link](#).

Este capítulo encerra a exposição teórico-prática dos estudos desenvolvidos nesta pesquisa. A partir daqui, passaremos às considerações finais, onde apresentaremos as conclusões a que chegamos, retomaremos as questões e hipóteses de pesquisa, assim como abordaremos as perspectivas que deste trabalho resultam.

11 CONCLUSÕES E PERSPECTIVAS

Ao chegarmos ao final deste trabalho, com a realização de alguns estudos-pilotos e depois com as análises empreendidas nos *corpora* em sua totalidade, cremos que temos elementos para responder às questões de pesquisa inicialmente postas.

Porém, antes de apresentarmos concretamente essas respostas, registramos as conclusões a que pudemos chegar com a comparação entre os resultados de cada Estudo-Piloto com os resultados das análises feitas com a totalidade dos *corpora*.

Estudo-Piloto 1 x Totalidade dos *Corpora*

No Estudo-Piloto 1, que representou uma primeira experiência nossa com a ferramenta Coh-Matrix-Dementia, chamou-nos muito a atenção o fato de um jornal popular ter figurado como um dos mais complexos entre os *corpora* analisados. Esse resultado foi alcançado com o somatório dos índices de todas as métricas analisadas naquele momento comparado com os escores do IF.

Ocorre que algumas métricas, como comentamos naquele Estudo-Piloto, foram redundantes e seus resultados pouco expressivos para serem aplicados na totalidade de nossos *corpora*. São elas: *cláusula por sentença*; *número de sentenças por parágrafo* e *número de palavras por parágrafo* e *sílaba por palavra de conteúdo*.

A primeira – *cláusula por sentença* –, em razão de apenas contar o número de verbos, sem distinguir, por exemplo, a presença de locuções verbais, não pôde ajudar na verificação do número de orações da frase. As três métricas seguintes – *número de sentenças por parágrafo*, *número de palavras por parágrafo* e *sílaba por palavra de conteúdo* –, conforme também comentamos nas análises a elas referentes, trazem dados que podem ser verificados em outras métricas, como o Índice Flesch.

Outra métrica utilizada nesse Estudo-Piloto 1 foi a relativa *ao número de palavras antes dos verbos principais*. Por essa métrica, o JM ficou como o mais difícil, o que nos surpreendeu por ser um jornal popular.

Utilizamos essa mesma métrica na comparação com a totalidade dos *corpora* e depois a descartamos. Assim fizemos porque, em nossas análises, nos demos conta de que muitas formas nominais dos verbos poderiam estar sendo interpretadas pela ferramenta como verbos, e isso poderia mascarar o resultado. Inclusive, trouxemos o caso do particípio “demandada” referindo-se, como substantivo, a uma das partes do processo. Este particípio foi alvo de análise

equivocada pelo Parser Palavras, conforme observamos no Estudo-Piloto 3. Diante disso, os resultados dessa métrica não puderam ser validados em nossa pesquisa.

Desse Estudo-Piloto 1, **sobraram as seguintes**: *incidência de negações, incidência de palavras de conteúdo, incidência de conectivos, referência anafórica adjacente e sobreposição de argumentos adjacentes*.

Dessas, as que **apresentaram escores que deixaram os jornais populares entre os mais complexos** foram: *incidência de negações, incidência de palavras de conteúdo e sobreposição de argumentos adjacentes*.

A métrica relativa à *negação* foi uma das métricas em que os jornais populares ficaram mais complexos no Estudo-Piloto 1. Na comparação feita com a totalidade dos *corpora*, conforme Capítulo 6, o resultado foi o oposto: os jornais populares ficaram como mais fáceis nesse quesito. Cremos que este resultado seja mais compatível com a realidade, uma vez que a comparação foi feita com um conjunto maior de textos do que a realizada no Estudo-Piloto 1.

Em relação à métrica *incidência de palavras de conteúdo*, o JM ficou como o segundo mais complexo, e o DG em quarto lugar. Na análise com a totalidade dos *corpora*, as palavras de conteúdo foram analisadas indiretamente na métrica *Densidade Semântica*. Esta métrica considera o número de palavras de conteúdo (lexicais – substantivo, adjetivo, advérbios e verbos) dividido pelo número de palavras funcionais (gramaticais – preposições, conjunções, numerais e artigos). Por essa métrica, os dois jornais populares continuaram como os mais difíceis.

Esse resultado é compatível com a TTR (relação entre *types* e *tokens*) que apresentaram. O DG foi o *corpus* que apresentou maior TTR, ou seja, que apresentou maior variação lexical, e o JM o segundo. Em nossas análises, ponderamos que essa maior variação e também maior densidade semântica podem não significar uma maior dificuldade, visto que palavras como **açougueiro, manicure** (com ocorrência nesses jornais) são bem mais fáceis de serem entendidas pelo público-alvo deste trabalho do que **tutela, homologação** (com ocorrência no CE), etc.

Pela métrica relativa à *sobreposição de argumentos adjacentes*, o DG ficou como o mais difícil, e o JM ficou mais fácil que os jornais ZH e FSP. A ordenação do mais difícil para o mais fácil, neste caso, foi: DG (0,169); ZH (0,395). FSP (0,398), JM (0,546) SP1 (1,056) e SP2 (1,186). Relembrando: quanto maior o índice, mais fácil é o texto, porque há mais repetição de palavras.

Na análise com a totalidade dos *corpora*, no Capítulo 8, o DG permaneceu como o mais difícil, e o JM oscilou do 3º colocado em relação aos referentes que se repetem nas sentenças

adjacentes e como o mais fácil em relação aos argumentos que se repetem entre as sentenças do texto. O resultado do DG é compatível com a maior TTR que apresenta. Tendo o texto mais variado lexicalmente, é esperado que tenha menor repetição de argumentos. O resultado do JM se mantém mais ou menos estável, até porque no estudo com a totalidade dos *corpora*, foram duas métricas relacionadas aos referentes.

De todas as métricas analisadas, então, no Estudo-Piloto 1, os jornais populares ficaram mais complexos apenas em relação à incidência de palavras de conteúdo e à sobreposição de argumentos adjacentes. Entretanto, como ponderamos acima, tais escores não refletem uma efetiva maior complexidade em relação aos demais *corpora* aqui analisados.

Estudo-Piloto 2 x Totalidade dos Corpora

No Estudo-Piloto 2, realizamos comparações entre os lemas do *corpus* de Sentenças utilizadas no Estudo-Piloto 1; do *corpus* do DG utilizado também no Estudo-Piloto 1; e do CorPop. Os dados desse Estudo-Piloto 2 revelaram que as sentenças e os textos do DG apresentam grande disparidade quanto ao vocabulário empregado. As sentenças apresentaram 73% de lemas que não ocorriam no DG. Relativamente ao CorPop, essa disparidade foi menor (38,09%), mas neste caso ponderamos que a lista do CorPop era bem maior que a lista de lemas do DG, o que gerava uma probabilidade maior de abarcamento do vocabulário das sentenças.

Da comparação entre o total de *types* do CE (14.665) e o total de *types* do *corpus* do DG, no Capítulo 9, verificamos que 6.797 não se encontram no DG. Isso gera um percentual de 46.34% de palavras das sentenças com alta probabilidade de não serem conhecidas dos leitores desse jornal. Lembramos que a comparação realizada no Estudo-Piloto 2 foi com os lemas, e não com os *types*. Essa pode ser uma das razões para, nesse Estudo-Piloto, o percentual ter sido maior.

Em relação ao CorPop, dos 14.665 *types* exclusivos das sentenças, 6.324 não constam no CorPop. Isso gera um percentual de 43,12% de palavras com alta probabilidade de não serem conhecidas pelo cidadão com letramento médio do país.

Os dados apresentados no Capítulo 9, com a totalidade dos *corpora* representam melhor a realidade do vocabulário desses textos do que os apresentados no Estudo-Piloto 2, com pequenas amostras de tamanhos também desproporcionais. De todo modo, as duas análises revelam o quanto o vocabulário empregado nas Sentenças está distante do que um cidadão brasileiro com médio letramento conhece.

No Estudo-Piloto 2, também apresentamos reescritas simplificadas de alguns trechos de algumas das palavras das sentenças ausentes nos outros *corpora*. No Capítulo 10, realizamos

outros exercícios de simplificação com trechos e com uma sentença completa, para termos também a noção de uma simplificação no seu todo. Para esses exercícios, sempre com base nas comparações lexicais que fizemos, procuramos partir de palavras das Sentenças que tivessem pouca ou nenhuma ocorrência nos *corpora* em contraste. Essa conduta partiu do pressuposto de que pouca ou nenhuma frequência implica pouco ou nenhum conhecimento dessa palavra por parte dos leitores dos *corpora* em contraste.

Os trechos originais e os reescritos foram medidos por alguns índices: no Estudo-Piloto 2, foram o IF, DS e TTR; no Capítulo 10, foram o IF e a Estatística Honoré. Em ambos, os índices revelaram uma facilitação do texto reescrito de modo simplificado. Conforme observamos, tais índices não são suficientes para aferir a totalidade dos aspectos relacionados à complexidade textual. Nesse sentido, as análises das métricas do Capítulo 8 dão conta de mais características envolvidas nessa aferição e revelaram que as Sentenças apresentam um vocabulário e uma sintaxe mais complexa que os demais *corpora* analisados.

Estudo-Piloto 3

No Estudo-Piloto 3, analisamos os constituintes com função de sujeito e de complementos verbais em uma amostra de 110 Sentenças de nosso CE, entrando em alguns aspectos sintáticos dessa amostra de sentenças. Por ter sido uma análise manual, como explicamos ao apresentarmos este Estudo-Piloto, não nos foi possível realizar o mesmo procedimento com a totalidade dos demais *corpora*. Trabalhamos com uma amostra de 25% da totalidade do CE, que, segundo apontamos naquele momento, pode ser considerada representativa dessa totalidade.

Em relação aos sujeitos, encontramos um percentual de 63% desses constituintes frasais que demandam maior sobrecarga na memória de trabalho do leitor. Nesse percentual, encontram-se:

- a) Sujeitos elípticos. Para serem conhecidos, dependem de um processo de referenciação situacional e textual. E esse processo, para um leitor não muito proficiente, pode ser um fator de complexidade.
- b) Sujeitos pós-verbais. Como mencionamos, muitas pesquisas sobre leitura, indicam que, por estarem fora da ordem canônica da língua – sujeito-verbo-objeto –, sujeitos pós-verbais dificultam o processamento da leitura, principalmente a leitores com pouca proficiência.

- c) Pronominais. Esse tipo de constituinte exige um processo de referenciação para ser compreendido. Se essa referenciação não for transparente, como citamos durante nossas análises, a demanda de trabalho na leitura é maior e conseqüentemente mais complexa.

Analizamos também a extensão desses constituintes com função de sujeito, porque a extensão das árvores de dependência sintática é uma das variáveis utilizadas na medição da complexidade sentencial. Além disso, conforme Fulgêncio e Liberato (2010), o que mais importa não é o comprimento da frase em si, mas dos seus constituintes, porque é neles que se encerra o significado.

O que verificamos foi a existência de constituintes que, ainda que não muito extensos em número de palavras, frequentemente são delimitados por algum tipo de oração encaixada. E a presença de orações encaixadas, principalmente quando inseridas entre os constituintes maiores, também é um complicador na leitura, porque o leitor precisa processar várias informações antes de chegar à principal.

Em relação aos complementos verbais, encontramos:

- a) 8.847 CFCVs em forma de sintagmas nominais;
- b) 1259 CFCVs em forma de sintagmas oracionais; e
- c) 734 CFCVs representados por pronomes.

Em princípio, a grande maioria desses constituintes se apresenta na forma de sintagmas nominais, mas o que observamos foi que o núcleo desses sintagmas normalmente é expandido por uma série de especificações. Essas especificações, pela natureza do tema, são importantes, mas, para efeito de clareza e simplicidade, poderiam ser melhor distribuídas no trecho. Reproduzimos aqui o exemplo (7.37).

(7.37) A controvérsia dos autos tem por cerne à (sic) discussão sobre a restituição dos valores pagos a título de taxa de corretagem realizado (sic) pela autora aos corretores Senhores FULANO DE TAL, BELTRANO DE TAL e SICRANO DE TAL, em decorrência de contrato de promessa de compra de um imóvel realizado entre a autora e a corré XXXXXXXXX, em que havia previsão da existência do pagamento dessa comissão.

Nesse exemplo, o núcleo desse CFCV de 59 palavras é **discussão**. Tal substantivo é especificado com o que vem após **sobre**, que é representado por uma sucessão de itens subordinados entre si. Se a frase for dividida em duas, já se percebe uma melhora.

Reescrita do exemplo (7.37):

A controvérsia dos autos tem por cerne a discussão sobre a restituição dos valores pagos a título de taxa de corretagem realizada pela autora aos corretores Senhores FULANO DE TAL, BELTRANO DE TAL e SICRANO DE TAL. O pagamento dessa comissão estava previsto no contrato de promessa de compra de um imóvel realizado entre a autora e a corré XXXXXXXXX.

Estudo-Piloto 4

No Estudo-Piloto 4, realizamos comparações entre os constituintes com função de sujeito e de complementos verbais em uma subamostra do CE. Ou seja, extraímos da amostra analisada no Estudo-Piloto 3 um conjunto de sentenças que compuseram a Amostra 2. Comparamos essa Amostra 2 com uma amostra de textos do DG e outra do JM. Tal como explicamos no Estudo-Piloto 3, não foi possível estender a comparação à totalidade dos *corpora*, mas julgamos que as amostras foram representativas, conforme várias vezes mencionamos no decorrer das análises.

Em relação aos sujeitos, a amostra das Sentenças apresentou:

- a) constituintes com maior número de palavras que os outros *corpora*,
- b) maior ocorrência de sujeitos elípticos, pós-verbais e oracionais.

Em relação a esses quesitos, então, as sentenças ficaram mais complexas que os jornais populares contrastados.

E as amostras dos jornais DG e JM apresentaram:

- a) sujeitos pré-verbais mais extensos que as sentenças.
- b) sujeitos pronominais em maior quantidade que nas sentenças (mais que o dobro).

Nesses quesitos, então, mostraram-se mais complexos que as sentenças.

Um fato que nos chamou muito a atenção foi **a ocorrência de sujeitos elípticos muito semelhante nos três corpora**. Ao longo das análises, explicamos as possíveis razões para essas ocorrências nas Sentenças. Quanto aos jornais, temos o estudo de Finatto *et al.* (2011), que analisou textos do DG e do jornal Zero Hora e chegou à conclusão de que a elipse não seria um fator decisivo na avaliação da complexidade textual naquele cenário discursivo. Mas quanto à frequência de sujeitos elípticos, foco do nosso trabalho aqui, podemos supor que as elipses seriam um dos recursos empregados para evitar a repetição de palavras.

Questionário

A aplicação do questionário, por sua vez, apesar do número reduzido de respondentes, primeiro nos permitiu validar os resultados das simplificações realizadas no Estudo-Piloto 2. As reescritas consideradas mais simples nesse Estudo-Piloto 2 foram também assim consideradas pelos respondentes do questionário. Validou também os resultados advindos das análises feitas por ferramentas computacionais utilizadas em nossa pesquisa, nos Estudos-Pilotos 1, 2 e 3.

Cumpre-nos retomar aqui a redução gradual dos respondentes à medida que o questionário avançava: de 99 para 66 ao final. Aventamos algumas hipóteses para esse fato: responder a um questionário sobre um tema que não se relaciona imediatamente com a realidade do respondente pode ser muito desagradável; questões que envolvam leitura atenta demandam mais tempo e concentração. De todo modo, este comportamento também pode nos indicar o quanto a linguagem jurídica está desconectada do cidadão brasileiro, que é o seu destinatário principal. E isso, por sua vez, é mais um fator a reforçar a necessidade de estudos como o que estamos empreendendo nesta pesquisa de doutorado.

Infelizmente, não pudemos dar prosseguimento à aplicação do questionário a mais respondentes no decorrer da pesquisa em razão da pandemia e em razão da atenção que tivemos de concentrar na análise dos dados relativos aos *corpora*. Como já registramos, muitos foram de forma manual, o que demandou muitas horas de trabalho. Além disso, durante a realização da pesquisa, também nos envolvemos com a produção do Guia de Linguagem Simples do TJRS, tarefa que igualmente nos tomou muitas horas de trabalho.

Diante disso, com base nesses resultados parciais e com as análises empreendidas com a totalidade dos *corpora*, conforme registramos nos Capítulos 8 e 9, podemos retomar as questões de pesquisa e suas hipóteses.

11.1 RETOMADA DAS QUESTÕES DE PESQUISA

Após as considerações acima, julgamos importante retomar as questões de pesquisa que motivaram nosso trabalho de modo a verificar se o que realizamos em nosso trabalho pode responder às perguntas que nos mobilizaram desde o início da trajetória desta tese.

As questões mais gerais de nossa pesquisa foram:

QUESTÃO 1

Do ponto de vista dos Estudos da Linguagem, do Texto, da Terminologia, da CT e ATT e da LS, as Sentenças dos JECs podem ser consideradas textos complexos para leitores adultos de escolaridade limitada?

QUESTÃO 2

Sendo as Sentenças dos JECs consideradas complexas para leitores adultos de escolaridade limitada, o que podem fazer, concretamente, os estudiosos da Linguagem, do Texto, da Terminologia, da CT e ATT e da LS para diminuir essa complexidade?

Para respondermos a ela com mais precisão, estabelecemos questões específicas que a seguir passamos a apresentar.

a) Como as Sentenças sob exame se configuram do ponto de vista lexical, morfossintático e terminológico?

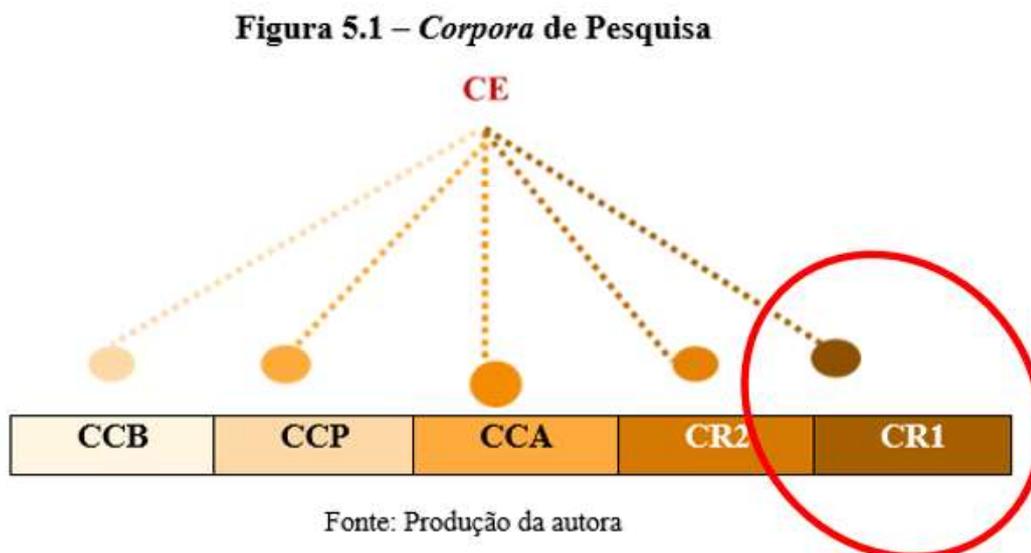
No que se refere ao vocabulário, os dados revelam que as Sentenças dos JECs, por serem textos especializados, apresentam muita terminologia da área jurídica e também muitas palavras “potencialmente complexas”. Como vimos, palavras que não figuram em textos que representam o PPE, veiculado nos jornais populares pesquisados e também no CorPop, são tidas como “potencialmente complexas”.

No Capítulo 9, em que comparamos os padrões lexicais dos *corpora* desta pesquisa, chegamos aos seguintes dados, em relação aos *types* exclusivos do CE:

- a) 39% não constam no jornal ZH;
- b) 38% não constam no jornal FSP;
- c) 45% não constam no jornal JM;
- d) 46 % não constam no jornal DG;
- e) 43% não constam no CorPop; e
- f) 34% não constam no CR1.

Esses percentuais indicam que **os leitores dos textos representados por esses *corpora* terão, no mínimo, 34% de chance de não compreender o vocabulário empregado nas Sentenças aqui analisadas**. E chamamos a atenção que o menor índice é relativo ao CR1 - *corpus* de obras literárias brasileiras clássicas. Pressupõe-se que o leitor dessas obras seja uma pessoa com mais letramento que a que tenha no máximo o Ensino Fundamental completo.

Pressupõe-se também que o vocabulário das Sentenças está mais próximo dos textos literários do que dos jornais, conforme Figura 5.1, reproduzida a seguir.

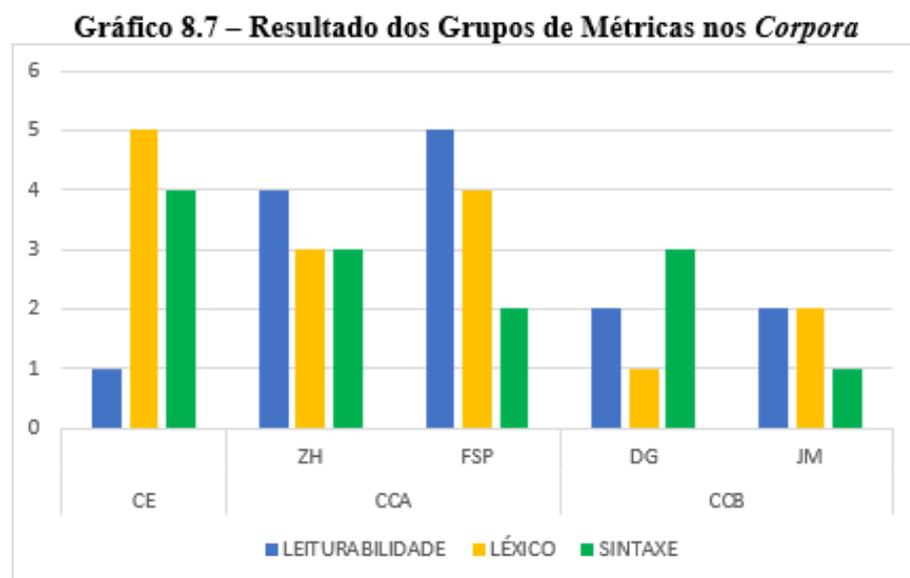


Quanto aos aspectos frasais, uma das características mais marcantes verificadas foi a extensão das frases das Sentenças, geralmente muito longas. Nos Estudos-Piloto 3 e 4, verificamos muitos constituintes frasais nem sempre tão extensos, porém a grande maioria complementados por orações encaixadas. Como já referimos, sentenças encaixadas, principalmente quando intercaladas, demandam mais trabalho do leitor e, conseqüentemente, aumentam a complexidade sentencial.

Além disso, como vimos no Estudo-Piloto 1 e também no Capítulo 8, as Sentenças apresentaram um índice de conectivos inferior ao dos demais *corpora*. Conforme explicamos nas análises dos dados, uma baixa incidência de conectivos torna o texto, em princípio, mais difícil, pois as relações lógicas não ficam expressamente estabelecidas. Essa baixa incidência de conectivos pode estar relacionada à presença de muitas orações subordinadas reduzidas (também muito comuns em textos jurídicos), que, em princípio, nem sempre são claras na relação que estabelecem, como ocorre frequentemente com o gerúndio. No Estudo-Piloto 3, 42% dos sujeitos elípticos encontrados têm seus verbos em formas nominais. E os sujeitos elípticos representam o tipo de sujeito de maior ocorrência na amostra de Sentenças analisadas.

No Capítulo 8, analisamos todos os *corpora* com base em algumas métricas do sistema NILC-Metrix (Coh-Metrix 3.0, de 2020). Essas métricas mediram fatores relacionados à

leiturabilidade, à sintaxe e ao vocabulário empregado nos *corpora*. E aqui reproduzimos o Gráfico 8.7, que sintetiza essas análises.



Relembrando: o eixo vertical apresenta a numeração por ordem de dificuldade dos *corpora*. O número 1 indica o texto mais fácil e o número 5 o mais difícil.

Tendo o CE como foco da análise aqui, este é o *corpus* mais complexo em relação ao léxico (coluna amarela) e à sintaxe (coluna verde).

b) Como se podem estimar eventuais níveis de dificuldade e/ou de complexidade linguístico-textual dessas Sentenças?

Os Estudos-Pilotos realizados e as análises dos Capítulos 8 e 9, considerando aspectos de leiturabilidade, vocabulário e sintaxe dos *corpora* em comparação nos indicam que esses níveis de dificuldade e/ou complexidade linguístico-textual das Sentenças podem ser verificados de várias formas:

- a) por meio de ferramentas computacionais, como as que utilizamos, que analisam conjuntos de textos por meio de várias métricas;
- b) por meio de contrastes com outros conjuntos de textos, como fizemos;
- c) por meio de análises manuais como as que realizamos nos Estudos-Piloto 3 e 4;
- d) por meio de entrevistas e/ou questionários com leitores reais, como o que relatamos na Seção 6.3.

c) Que elementos lexicais, morfossintáticos e terminológicos podem contribuir para que os requisitos legais de simplicidade e informalidade expressos no art. 2º da Lei n. 9.099, de 1995, nas Sentenças sob exame, não sejam atendidos?

Entre os elementos lexicais, terminológicos e frasais que podem contribuir para que os requisitos de simplicidade e informalidade, nas Sentenças sob exame, não sejam atendidos, estão:

- a) presença de palavras “potencialmente complexas”, segundo as comparações realizadas no Estudo-Piloto 2 e no Capítulo 9;
- b) presença de muita terminologia;
- c) frases muito longas com vários níveis de hierarquia sintática;
- d) constituintes da oração (sujeitos elípticos, pronominais e pós-verbais; sujeitos e complementos verbais expandidos por muitas orações encaixadas) que demandam maior sobrecarga na memória de trabalho do leitor;
- e) pouca presença de conectivos;
- f) muita incidência de negação.

d) Quais as melhores estratégias de escrita e/ou reescrita para que os textos das Sentenças sob exame possam atender seus requisitos legais de informalidade e simplicidade?

Separamos um capítulo deste trabalho para sugestões de simplificação que, certamente, contribuirão para que os requisitos legais de informalidade e simplicidade das Sentenças dos JECs sejam atendidos. De todo modo listamos aqui algumas estratégias de escrita e/ou reescrita para que os textos possam atender a essa finalidade.

- a) substituição das palavras “potencialmente complexas” por sinônimos mais simples. Esses sinônimos podem ser extraídos do CorPop, já disponível na internet, assim como de dicionários de sinônimos.

Exemplo: **versa** por **trata** em:

De	Para
A lide versa a respeito de defeito do serviço.	A lide trata a respeito de defeito do serviço.

- b) substituição de terminologias, quando possível, por palavras mais simples.

Exemplo: **lide** por **discussão** em:

De	Para
A lide trata a respeito de defeito do serviço.	A discussão trata a respeito de defeito do serviço.

- c) acréscimo de explicação quando a substituição por um sinônimo não for suficiente para contemplar a totalidade do significado da terminologia. É o que fizemos, no Capítulo 10, em uma das reescritas apresentadas para um trecho em que constava os termos **improcedência** e **demanda**. Repetimos o exemplo aqui:

De	Para
É o caso (1) de improcedência (2) da demanda (3).	É o caso de (1) improcedência (o pedido não tem previsão na lei ou o autor não provou o que pede) (2) da demanda (pedido) (3).

- d) diminuição do tamanho das frases, conforme exemplo (10.1), reproduzido aqui.

De	Para
Contudo, creio que tal <u>cláusula</u> não serve para afastar o direito da autora na medida em que não foi <u>redigida</u> de forma clara, pois não há prova de que a ré tenha esclarecido à consumidora expressamente que a cirurgia <u>em exame</u> estaria <u>excluída</u> de atendimento do plano de saúde, sendo que o capítulo que trata das <u>exclusões</u> do atendimento são extremamente obscuras (sic), já que não esclarecem quais seriam os procedimentos previstos na referida Tabela INAMP/SUS, não havendo prova de que a consumidora <u>teve vista</u> sobre ela, razão pela qual não se vincula a tal <u>exclusão contratual</u> .	Contudo , creio que tal cláusula não afasta o direito da autora porque não foi redigida de forma clara. Não há prova de que a ré tenha esclarecido a consumidora de que o plano de saúde não cobria a cirurgia em exame. O capítulo que trata das exclusões do atendimento é confuso, pois não esclarece quais os procedimentos previstos na referida Tabela INAMP/SUS. Além disso , não há prova de que a consumidora teve conhecimento dessa tabela. Assim , essa tabela não pode ser vinculada à referida exclusão contratual
Observação: A reescrita apresentada não alterou as palavras potencialmente complexas (sublinhadas) no trecho. Para uma simplificação mais abrangente, este processo também deverá ser observado.	

- e) diminuição do número de orações encaixadas nos constituintes da oração. Trazemos exemplo (5) do Estudo-Piloto 3:

De	Para
A questão [pela qual se insurge a autora], (sic) <u>tem</u> relação com a necessidade de uso dos dados pessoais de seu cartão de crédito[...]	A autora questiona a necessidade de uso dos dados pessoais do seu cartão crédito[...]

- f) diminuição de orações subordinadas reduzidas e inserção de mais conectivos para deixar as relações lógicas estabelecidas de forma mais clara. Trazemos o exemplo (15) do Estudo-Piloto:

De	Para
Restou comprovado nos autos a culpa exclusiva do demandado, que invadiu a via preferencial pela qual a demandante trafegava, desrespeitando a sinalização e a preferencialidade da via [...]	Restou comprovado nos autos a culpa exclusiva do demandado. O demandado INVADIU a via preferencial onde a demandante trafegava E DESRESPEITOU a sinalização e a preferencialidade da via [...]
Observação: Para maior clareza, a expressão o demandado é repetida numa nova frase introduzida por nós, e a subordinada introduzida pelo gerúndio desrespeitando foi substituída por uma coordenada aditiva com o verbo numa forma finita em concordância com o sujeito “o demandado”. A alteração proposta é apenas em relação ao tópico do item em questão.	

- g) redução da incidência de negação

Como vimos no Capítulo 8, *a incidência de negação* foi uma das métricas em que o CE ficou como mais complexo. E aqui observamos que a terminologia da área carrega consigo essa complexidade. É o caso da expressão **ação declaratória de inexistência de débito** no seguinte trecho.

De	Para
Embora a ação tenha sido nominada de ação declaratória de inexistência de débito, nem a descrição dos fatos e fundamentação negaram a existência do débito, limitando-se a questionar informações alegadamente contraditórias que teriam vindo da ré, nem foi formulado pedido de afastamento do valor cobrado.	<u>O nome da</u> ação é <i>ação declaratória de inexistência de débito</i> . <u>Ou seja, é</u> uma ação que pede que se declare que não existe dívida . <u>Apesar desse nome</u> dado à ação, a descrição e a fundamentação dos fatos negam que existe a dívida. <u>Apenas questionam</u> informações vindas da ré e que consideram contraditórias. <u>Além disso,</u> o pedido de afastamento do valor não foi feito.
Observação: Como se vê na reescrita apresentada, não substituímos o nome da ação, mas acrescentamos uma frase explicando o que tal nome significa. Além disso, para que as ideias do trecho original ficassem mais ordenadas e, conseqüentemente, de mais simples leitura, realizamos várias alterações. Entre essas alterações está a transformação de 1 frase só em 5, e a substituição de débito por dívida .	

Com base nas respostas às questões acima, cremos que as questões gerais de nossa pesquisa também tenham sido respondidas. Feito isso, passamos a confrontar os resultados de nossas análises com o conjunto de hipóteses que estabelecemos.

11.2 RETOMADA DAS HIPÓTESES DE PESQUISA

Considerando as questões anteriores, estabelecemos um conjunto de hipóteses gerais de trabalho.

HIPÓTESE 1

O texto da maioria das Sentenças dos JECs do PJRS apresenta padrão lexical, morfossintático e terminológico pouco coincidente com padrões do Português Popular Escrito (PPE). Tais características podem dificultar sua inteligibilidade para pessoas de escolaridade limitada e pouca experiência em leitura.

Os resultados das análises empreendidas em todos os estudos, em especial no Estudo-Piloto 2 e no Capítulo 9, permitem concluir que os textos das Sentenças analisadas tendem a apresentar padrões lexicais pouco coincidentes com os padrões do PPE, pela presença de muitas palavras “potencialmente complexas” e de terminologias. E isso pode dificultar a compreensão por parte de pessoas de escolaridade limitada e pouca experiência em leitura.

Nos Estudos-Piloto 3 e 4, com a análise dos constituintes frasais com função de sujeito e de complementos verbais em amostras de Sentenças do CE, ficou evidenciada a complexidade desses constituintes: 63% dos sujeitos analisados apresentam-se em tipos (elípticos, pronominais e pós-verbais) que demandam maior sobrecarga na memória de trabalho do leitor. Além desse dado, a grande maioria dos constituintes frasais, tanto sujeitos como complementos verbais, apresenta-se delimitada por orações encaixadas que, como vimos, também demandam mais do leitor e, conseqüentemente, são um dificultador para a compreensão leitora.

Desses resultados, o que fica demonstrado é que não só o padrão lexical, mas também a sintaxe das orações das Sentenças dos JECs se apresenta de modo pouco coincidente com o do PPE.

HIPÓTESE 2

Os resultados das pesquisas relativas aos estudos do Texto, da Linguagem e da Terminologia podem subsidiar a elaboração de propostas de escritas das Sentenças dos

JECs do PJRS que promovam maior acessibilidade textual e terminológica a leitores adultos de escolaridade limitada.

Os resultados das simplificações efetuadas no Estudo-Piloto 2 e no Capítulo 10 confirmam esta hipótese. Pelas simplificações efetuadas e também pelas respostas ao questionário, foi possível verificar que os resultados das pesquisas relativas aos estudos do Texto, da Linguagem e da Terminologia podem subsidiar a elaboração de propostas de escritas das Sentenças dos JECs do PJRS que promovam maior acessibilidade textual e terminológica a leitores adultos de escolaridade limitada.

E, em relação às questões específicas, as hipóteses que estabelecemos foram as seguintes:

Hipótese a

O estudo dos padrões lexicais, morfossintáticos e terminológicos pode auxiliar na caracterização da configuração lexical, morfossintática e terminológica dos textos das Sentenças dos JECs do PJRS.

Considerando o que até aqui expusemos, podemos afirmar que as análises empreendidas neste trabalho foram fundamentais para melhor caracterizar a configuração lexical, morfossintática e terminológica dos textos das Sentenças dos JECs do PJRS.

Nas análises relativas aos constituintes frasais, foi possível dimensionar suas características principais. Nas análises lexicais, foi possível constatar muitas palavras e terminologias que ocorrem somente nessas Sentenças.

Hipótese b

A comparação dos textos das Sentenças dos JECs do PJRS com textos que representem o PPE e também com textos de literatura dos sécs. XIX e XX pode ajudar na verificação de eventuais níveis de complexidade linguístico-textual e terminológica dessas Sentenças.

Os resultados das comparações realizadas neste trabalho ajudaram, e muito, a verificar eventuais níveis de complexidade linguístico-textual das Sentenças aqui analisadas. Nas considerações iniciais deste capítulo e nas respostas às questões específicas acima, essa situação fica bem assentada.

Como se pôde observar, as Sentenças aproximam-se de modo significativo dos padrões da literatura dos sécs. XIX e XX. E isso se revela como um estilo cultivado pelos redatores, conforme uma tradição discursiva.

Hipótese c

A presença de palavras e construções sintáticas não muito frequentes nos textos do PPE bem como a presença de muitos termos técnicos podem contribuir para que os requisitos da simplicidade e informalidade expressos no art. 2º da Lei n. 9.099, de 1995, não sejam atendidos.

O que se afirma nesta hipótese perpassa todo o trabalho realizado nesta tese. Durante as análises e nas propostas de reescrita, ficou patente que a presença de palavras “potencialmente complexas” organizadas numa sintaxe também complexa são fatores que efetivamente contribuem para que os requisitos da simplicidade e informalidade expressos no art. 2º Lei n. 9.099, de 1995, não sejam atendidos no que diz respeito à linguagem.

É o caso do seguinte trecho, extraído da sentença apresentada no Capítulo 10 deste trabalho: *No mérito, a análise do cotejo probatório carreado aos autos conduz à improcedência por parte autora do pedido articulado na peça preambular.*

Hipótese d

O emprego de palavras e de uma sintaxe mais próximas das que se veem nos textos do PPE e a diminuição do uso de termos técnicos podem representar estratégias para simplificar o texto das Sentenças sob exame. Com isso, se atenderia aos requisitos da simplicidade e informalidade expressos no art. 2º da Lei n. 9.099, de 1995.

Creemos que as propostas de simplificação apresentadas ao longo do trabalho, em especial no Estudo-Piloto 2, no Capítulo 10 e nas respostas à Questão Específica **d** confirmam esta hipótese.

É o caso da reescrita simplificada que apresentamos ao trecho dado como exemplo para a hipótese anterior: *No mérito, a análise das provas trazidas aos autos não permite que o pedido inicial seja atendido.*

Enfim, cremos que todas as hipóteses estabelecidas neste trabalho foram confirmadas e avalizam o título deste trabalho: *Sentenças judiciais e Linguagem Simples: um encontro possível e necessário.*

Esse encontro mais do que possível, é extremamente necessário. Para tanto, outras iniciativas podem ser tomadas. A seguir, apresentamos algumas delas quando trazemos algumas perspectivas que este trabalho apresenta.

11.3 PERSPECTIVAS

Este trabalho evidenciou que é possível a redação de uma sentença judicial com uma linguagem mais simples e mais próxima do cidadão que recorre à Justiça para resolver algum conflito. E, conforme mencionamos, hoje tem havido várias iniciativas nesse sentido no Brasil. Uma delas é o emprego das técnicas da *Visual Law*, que objetiva descomplicar a linguagem jurídica com a aplicação de conceitos visuais em documentos jurídicos como contratos, petições, sentenças, etc. A ideia é ir além de um *design* gráfico, mas também deixar o texto mais simples.

A seguir, na Figura 11.1 apresentamos uma sentença judicial da Justiça do Trabalho redigida com os princípios da *Visual Law*.

Figura 11.1 – Imagem de uma sentença da Justiça do Trabalho em *Visual Law*


 PODER JUDICIÁRIO
 JUSTIÇA DO TRABALHO
 TRIBUNAL REGIONAL DO TRABALHO DA 7ª REGIÃO

LEGAL DESIGN | VISUAL LAW
 O presente arquivo tem apenas o intuito de facilitar o acesso e a compreensão acerca da decisão proferida e de proporcionar às partes interessadas uma prestação judicial mais humanizada.

RESUMO DA SENTENÇA

ATSum

RECLAMANTE:

RECLAMADO:

RELATÓRIO/RESUMO DO PROCESSO

Dispensado, com fundamento no disposto no art. 952-A, da CLT.

FUNDAMENTAÇÃO:

PETIÇÃO INICIAL	DEFESA
A parte Reclamante pretende a condenação da parte Reclamada ao pagamento de: <ol style="list-style-type: none"> Vale-transporte; Férias em dobro + 1/3 dos períodos aquisitivos de 2014/2015, 2015/2016 e 2016/2017; Horas extraordinárias com reflexos; Benefício de gratuidade judicial. 	A parte Reclamada alega: <ol style="list-style-type: none"> Reclamante optou por não recebimento de vale-transporte; Férias foram gozadas e pagas; Gozou de intervalo remunerado.

AUDIÊNCIA DE INSTRUÇÃO
A parte Reclamante confirmou o recebimento dos vídeos das contracheiras. Não foram produzidos provas orais.

DISPOSITIVO

Pelo exposto e tudo o mais que dos autos conste, nos termos da fundamentação, julgo **IMPROCEDENTES** os pedidos constantes na reclamatória em epígrafe, salvo o de gratuidade judicial, que ora defiro.

Sucumbente a parte Reclamante, arbitro **honorários advocatícios** em favor dos patronos da parte Reclamada no percentual de 5% do valor atribuído à demanda, na forma do art. 791-A, da CLT.

Custas pelo Reclamante, **no importe** de R\$786,17, calculadas sobre o valor da causa, dispensadas em face da gratuidade judicial deferida.

Intimem-se a parte Reclamante, na pessoa de seus patronos, através do DEJT.

Juiz do trabalho Substituto

LEGAL DESIGN | VISUAL LAW
 O presente arquivo de sentença tem apenas o intuito de facilitar o acesso e a compreensão acerca da decisão e de proporcionar às partes interessadas uma prestação judicial mais humanizada.


Sucumbente a parte Reclamante, arbitro **honorários advocatícios** em favor dos patronos da parte Reclamada no percentual de 5% do valor atribuído à demanda, na forma do art. 791-A, da CLT.

DISPOSITIVO

Pelo exposto e tudo o mais que dos autos conste, nos termos da fundamentação, julgo **IMPROCEDENTES** os pedidos constantes na reclamatória em epígrafe, salvo o de gratuidade judicial, que ora defiro.


Sucumbente a parte Reclamante, arbitro **honorários advocatícios** em favor dos patronos da parte Reclamada no percentual de 5% do valor atribuído à demanda, na forma do art. 791-A, da CLT.

Custas pelo Reclamante, no importe de R\$786,17, calculadas sobre o valor da causa, dispensadas em face da gratuidade judicial deferida.

Intimem-se a parte Reclamante, na pessoa de seus patronos, através do DEJT.

Fonte: <https://bernardodeazevedo.com/wp-content/uploads/2021/05/13-VT-fortaleza-resumo-sentenca.pdf> - Acesso em: 11 maio 2022.

O visual gráfico da sentença da Figura 11.1 é realmente atrativo e explicativo. O destaque para cada parte do texto da sentença sinaliza melhor ao leitor cada uma dessas partes e lhe dá possibilidade de ir, de imediato, ao que lhe interessa.

Como se pode ver, logo abaixo da sentença, inserimos um quadro em que repetimos as informações que constam na parte destacada pela moldura em vermelho na sentença. E, nesse quadro, sinalizamos algumas palavras que nos parecem ainda “potencialmente complexas” ao leitor-alvo do nosso trabalho aqui: um cidadão brasileiro adulto com escolaridade limitada e pouca experiência em leitura.

A primeira dessas palavras é **sucumbente**, que provém de **sucumbência**. Como vimos em nossas análises, **sucumbência**, além de ocorrer nas Sentenças aqui analisadas, ocorreu apenas 1 vez no jornal ZH. Nos demais *corpora*, inclusive no de obras literárias, não há registro dessa palavra. Trata-se de um **termo jurídico** que precisaria, para maior compreensão do texto, de uma explicação do que significa.

Ou seja, ainda que o *design* ajude, e muito, na compreensão do texto da sentença em questão, o vocabulário ainda precisaria de maiores ajustes. Essa, então, seria uma das perspectivas da tese que ora apresentamos: prosseguir com pesquisas relativas aos aspectos linguístico-textuais das Sentenças judiciais em comunhão de esforços com outras iniciativas, como as da *Visual Law*.

É preciso considerar também que as desigualdades sociais existentes em nosso país foram acirradas pela pandemia do Coronavírus. Vivemos um momento em que medidas que deem o peixe se impõem. Nesse sentido, para que as instituições governamentais e os tribunais realmente **garantam** o acesso à informação para a grande maioria dos brasileiros, uma das possibilidades é criar, em seus quadros funcionais, setores com profissionais especializados em procedimentos de simplificação da linguagem.

A seguir apresentamos algumas outras perspectivas que o presente trabalho apresenta:

- a) A expansão das comparações aqui realizadas, como, por exemplo:
- *comparar os constituintes frasais das sentenças com os dos textos literários;*
 - *comparar os constituintes frasais das sentenças com os dos jornais de referência aqui analisados.*

Essas comparações, tal como procedemos com as comparações lexicais, ajudariam a identificar melhor os padrões sintáticos das sentenças. Ou seja, seria possível identificar com mais precisão se estariam mais próximos de um português culto atual, como os empregados nos

jornais de referência, ou de um português mais erudito e literário como os das obras da literatura que compuseram o CR1 de nosso trabalho.

- *comparar as Sentenças dos JECs com Sentenças de outras esferas da Justiça.*

Como, nas demais esferas do Judiciário, é necessária a intermediação de um advogado, haveria, em princípio, uma menor preocupação com a linguagem, uma vez que este profissional serviria de intérprete para o cidadão. Não podemos esquecer, porém, que, no Brasil, de acordo com o Princípio da Obrigatoriedade das Leis expresso no art. 3º da Lei de Introdução às Normas do Direito Brasileiro (LINDB), ninguém pode ser desculpado por não cumprir a lei alegando que não a conhece. Mesmo que assim seja,

a própria lei recorre a mecanismos que impedem seus destinatários de apreendê-la. Trata-se, portanto, de um dos mais discutidos paradoxos da cultura legal contemporânea. É como se a prática legal e a linguagem jurídica se encontrassem estruturadas de tal forma que inviabilizam a aquisição desse conhecimento por qualquer pessoa que não pertença a uma “elite” de especialistas altamente treinados nos vários ramos do domínio jurídico (SANTANA, 2017).

A ideia, então, com essa comparação seria verificar em que medida a linguagem das Sentenças dos JECs seria efetivamente mais simples do que as de outras esferas da Justiça.

- b) A ampliação da aplicação do questionário que realizamos limitadamente nesta tese.

A retomada de um questionário a leitores reais com os dados advindos de nossas análises será de grande valia para a legitimação das simplificações oferecidas para trechos potencialmente complexos.

Salientamos que as análises aqui empreendidas, com exceção das que fizemos manualmente, só foram possíveis graças aos recursos e ferramentas criadas no âmbito da LCOMP ou Processamento da Linguagem Natural (PLN). A parceria entre linguistas e profissionais da computação é um caminho sem volta. Os linguistas, além de se valerem das informações extraídas pelas ferramentas do PLN, podem contribuir “para a melhoria desses sistemas, com a consideração de regras linguísticas que possam ser a ele integradas” (FINATTO; LOPES; CIULLA, 2015, p. 47).

Salientamos também a importância dos subsídios da Linguística de *Corpus* e dos estudos estatístico-lexicais. As comparações quantitativas realizadas neste trabalho evidenciam que tais referenciais teóricos não se referem a uma “mera contagem de palavras e outras construções sintáticas”. Os números resultantes das análises quantitativas são a base para uma

análise qualitativa real e qualificada. E aqui reproduzimos as palavras de Pasqualini, que resumem bem essa ideia e também a de que um *corpus* de textos é vinculado a pessoas reais:

os textos com os quais trabalhamos não saem do vácuo nem a ele se dirigem. Repetimos à exaustão que somos mais do que contadores de palavras, já que esta é uma crítica recorrente às nossas pesquisas que envolvem Linguística de *Corpus*. [...] Um *corpus* é mais do que somente uma sacola de textos e palavras. (PASQUALINI, 2018, p. 135)

Por fim, encerramos nossas considerações finais com a esperança de que este trabalho contribua efetivamente para diminuir a distância entre o cidadão e a Justiça. Para tanto, deixamos um poema que reflete sobre a justiça boa e a justiça ruim. Que o encontro entre as Sentenças judiciais e a Linguagem Simples produza no Brasil uma justiça boa.

A justiça é o pão do povo.
 Às vezes bastante, às vezes pouca.
 Às vezes de gosto bom, às vezes de gosto ruim.
 Quando o pão é pouco, há fome.
 Quando o pão é ruim, há descontentamento.

Fora com a justiça ruim!
 Cozida sem amor, amassada sem saber!
 A justiça sem amor, cuja casca é cinzenta!
 A justiça de ontem, que chega tarde demais!
 Quando o pão é bom e bastante
 O resto da refeição pode ser perdoado.
 Não pode haver logo tudo em abundância.
 Alimentado do pão da justiça
 Pode ser feito o trabalho
 De que resulta a abundância.

Como é necessário o pão diário
 É necessária a justiça diária.

Sim, mesmo várias vezes ao dia.
 De manhã, à noite, no trabalho, no prazer.
 No trabalho que é prazer.

Nos tempos duros e nos felizes
O povo necessita de pão diário
Da justiça, bastante e saudável.

Sendo o pão da justiça tão importante
Quem, amigos, deve prepará-lo?

Quem prepara o outro pão?

Assim como o outro pão
Deve o pão da justiça
Ser preparado pelo povo.

Bastante, saudável, diário.

Bertolt Brecht, dramaturgo alemão (1898-1956),
na tradução de Paulo César Lima de Souza (1986).

REFERÊNCIAS

- ADELSTEIN, Andreína. Representación de Unidades Terminológicas: el modelo de término wüsteriano. **Organon**, Porto Alegre, v. 12, n. 26, p.66-81, 1998.
- AGUIAR JÚNIOR, Ruy Rosado de. *Ementas e sua técnica*. **Revista de Doutrina da 4ª Região**, Porto Alegre, n. 27, dez. 2008. Disponível em: http://www.revistadoutrina.trf4.jus.br/artigos/edicao027/ruy_rosado.html Acesso em: 13 abr. 2017.
- ALMEIDA, G. M. B. A Teoria Comunicativa da Terminologia e a sua prática. **Alfa**, São Paulo, 50 (2): 85-101, 2006.
- ALUÍSIO, Sandra Maria; ALMEIDA, Gladis Maria de Barcellos. O que é e como se constrói um *corpus*?: Lições aprendidas na compilação de vários *corpora* para pesquisa linguística. **Calidoscópico**, São Leopoldo, v. 4, n. 3, p.156-178, set./dez, 2006.
- AMARAL, M. F. **Jornalismo Popular**. São Paulo: Contexto, 2006.
- ANDRIGHI, Fátima Nancy. **Pela compreensão da justiça**. Brasília, DF, 2005. Disponível em: <http://bdjur.stj.jus.br/dspace/handle/2011/1764>. Acesso em: 10 fev. 2017.
- ASSOCIAÇÃO DOS MAGISTRADOS BRASILEIROS – AMB. **O Judiciário ao Alcance de Todos** – Noções Básicas de Juridiquês. Disponível em: <http://www.amb.com.br/portal/juridiques/livro.pdf>. Acesso em: 07 mar. 2010.
- BAILIN, A.; GRAFSTEIN, A. **Readability: Text and Context**. London: Palgrave Macmillan, 2016.
- BAKER, Eva. L.; ATWOOD, Nancy K.; DUFFY, Thomas M. Cognitive Approaches to Assessing the Readability. In: DAVISON, Alice; GREEN, Georgia M. (eds.) **Linguistic complexity and text comprehension. Readability issues reconsidered**. Hillsdale, N.J.: Lawrence Erlbaum Associates, 1988.
- BARROS, Diana Luz Pessoa de. **Teoria Semiótica do Texto**. São Paulo: Ática. 2005.
- BEILKE, Neubiana Silva Veloso. **Pommersche Korpora**: uma proposta metodológica para compilação de *corpora* dialetais. 2016. 285 f. Dissertação (Mestrado) - Curso de Letras, Programa de Pós-Graduação em Estudos Linguísticos., Universidade Federal de Uberlândia, Uberlândia, 2016.
- BERBER SARDINHA, Tony. **Linguística de Corpus**. Barueri: Manole, 2004.
- BERBER SARDINHA, Tony. Linguística de *Corpus*: histórico e problemática. **D.E.L.T.A.**, vol.16, n. 2 :323-367, 2000.
- BERTI, Fábio Ramos. **Projeto Simplifica traduz termos jurídicos para a população**. 2020. Disponível em: <https://www.tjrs.jus.br/novo/noticia/projeto-simplifica-traduz-termos-juridicos-para-a-populacao/>. Acesso em: 12 fev.2021.

BIBER, Douglas. Representatividade em planejamento de *corpus*. In: FINATTO, M.J.B.; ZILIO, L.; GONÇALVES, F. (Orgs.). Traduzido por Suzana Gonçalves e Fabiano Gonçalves. **Cadernos de Tradução** (Porto Alegre), v. 30, UFRGS, 2012.

BIBER, Douglas. **Variation across Speech and Writing**. Cambridge: Cambridge University Press, 1988.

BIBER, Douglas. **Dimensions of Register Variation – A Cross-Linguistic Comparison**. Cambridge: Cambridge University Press, 1995.

BICK, E. **The Parsing System PALAVRAS**: automatic grammatical analysis of Portuguese in a constraint grammar framework. Aarhus: Aarhus University Press, 2000.

BIDERMAN, Maria Tereza Camargo. Léxico e Vocabulário Fundamental. **Alfa**, São Paulo, 40: 27-46, 1996.

BIDERMAN, M. T. A face quantitativa da linguagem: um dicionário de frequências do português. **Alfa**, São Paulo, v. 42, número especial, p. 161-181, 1998.

BIDERMAN, M. T. C. Estatística linguística. **Alfa**, São Paulo, v. 11, p. 117- 128, 1967.

BOULANGER, Jean-Claude. Présentation: images et parcours de la socioterminologia. **Meta**, vol. 40, n. 2, 1995, p. 194-205.

BRASIL. Lei nº 10.406, de 10 de janeiro de 2002. **Código Civil**.

BRANCO, Sara Isabel Antunes. **A Compreensão da Terminologia Legal por Adultos**. 2014. 116 f. Dissertação (Mestrado) - Curso de Instituto de Ciências Biomédicas Abel Salazar, Universidade do Porto, Porto, 2014. Disponível em: <https://repositorio-aberto.up.pt/bitstream/10216/77789/2/33833.pdf>. Acesso em: 15 mar. 2017.

BRASIL. Lei nº 9.099, de 26 de setembro de 1995. **Lei dos Juizados Especiais**.

BRASIL. MINISTÉRIO DA JUSTIÇA. **Atlas de Acesso à Justiça**: Glossário. Disponível em: <http://www.acessoajustica.gov.br/pub/sobre/oAtlas/exibirOAtlas.faces>. Acesso em: 20 out. 2016.

BRASIL. Presidência da República. **Manual de Redação da Presidência da República**. 3. ed. Brasília: Presidência da República, 2018. Disponível em: <http://www4.planalto.gov.br/centrodeestudos/assuntos/manual-de-redacao-da-presidencia-da-republica/manual-de-redacao.pdf>. Acesso em: 12 maio 2022.

BRECHT, Bertold. **Poemas 1913-1956**. Seleção e Tradução de Paulo César Lima de Souza. São Paulo: Editora 34, 2012, 360 p.

CABRÉ, Maria Teresa. **La terminologia**. Barcelona: Ed. Antártida. 1999.

CARPIO, P. M. S. **Abaixando o cocho: adaptação de textos sobre doenças causadas pela inalação de amianto destinados para o público leigo.** Trabalho de conclusão de curso. Bacharelado em Letras. Porto Alegre, UFRGS, 2017.

CARVALHO, Adilson. **Linguagem jurídica - Uma porta (fechada) para o acesso à justiça.** Disponível em: <https://jf-ms.jusbrasil.com.br/noticias/140750/linguagem-juridica-uma-porta-fechada-para-o-acesso-a-justica>. Acesso em: 13 fev. 2017.

CAVALHEIRO, Patrícia da Cruz. **DESCOMPLICA: Comissão de Inovação do TJRS lança projeto para simplificar o texto jurídico.** 2020. Disponível em: <<https://www.tjrs.jus.br/novo/noticia/descomplica-comissao-de-inovacao-do-tjrs-lanca-projeto-para-mudar-o-texto-juridico/>>. Acesso em: 12 fev. 2021.

CHOMSKY, Noam. **Current Issues in Linguistic Theory.** Paris : Mouton, 1970.

CIAPUSCIO, G. E. La terminología desde el punto de vista textual: selección, tratamiento y variación. **Organon**, Porto Alegre, v. 12, n. 26, p. 43-65, 1998.

CIAPUSCIO, G. E. **Textos especializados y terminologia.** Barcelona: IULA, 2003.

COLARES, Virgínia (Org.). **Linguagem e direito.** Recife: UFPE, 2010. 338 p. (v. 500). Disponível em: <ler.letras.up.pt/uploads/ficheiros/12697.pdf>. Acesso em: 10 fev. 2017.

CORNU, Gérard. (1990) **Linguistique juridique.** Paris: Monchrestien, 1990, 412p.

COSCARELLI, C. V. Entendendo a Leitura. **Estudos da Linguagem**, Belo Horizonte, v.10, n.1, p.7-27, jan./jun. 2002.

CUNHA, André Luiz Verucci da. **Coh-Metrix Dementia: análise automática de distúrbios de linguagem nas demências utilizando processamento de línguas naturais.** 2015. 155 f. Dissertação (Mestrado) - Curso de Ciências da Computação, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2015.

CUNHA, Rogério Sanches. **Lei 13.603/18:** Acrescenta a simplicidade nos processos dos Juizados Especiais Criminais. Disponível em: <https://meusitejuridico.editorajuspodivm.com.br/2018/01/10/lei-13-60318-acrescenta-simplicidade-nos-processos-dos-juizados-especiais-criminais/>. Acesso em 15 out. 2019.

DAVISON, Alice; GREEN, Georgia M. (eds.) **Linguistic complexity and text comprehension. Readability issues reconsidered.** Hillsdale, N.J.: Lawrence Erlbaum Associates, 1988.

DELL'ORLETTA, F.; WIELING, M.; CIMINO, A.; VENTURI, G.; MONTEMAGNI, S. Assessing the readability of sentences: Which *corpora* and features? **Proceedings of the Ninth Workshop on Innovative Use of NLP for Building Educational Applications**, p. 163–173, 2014.

DELL'ORLETTA, F.; MONTEMAGNI, S.; VENTURI, G. Read-it: Assessing readability of italian texts with a view to text simplification. **Proceedings of the 2nd Workshop on Speech and Language Processing for Assistive Technologies**, p. 73–83, 2011.

DUBAY, W. H. **The Principles of Readability**. 25 August 2004. Disponível em: <http://www.impact-information.com/impactinfo/readability02.pdf>. Acesso em: 13 jul. 2017.

DUARTE, Elisa Teixeira. **A Lingüística de *Corpus* a serviço do tradutor**:: proposta de um dicionário de culinária voltado para a produção textual. 2008. 400 f. Tese (Doutorado) - Curso de Letras, Programa de Pós-Graduação em Estudos Lingüísticos e Literários em Inglês do Departamento de Letras Modernas, Universidade de São Paulo., São Paulo, 2008.

EVERS, Aline. **A redação engaiolada**: padrões lexicais e ensino de redação em cursos pré-vestibulares populares. 2018. 229 f. Tese (Doutorado em Letras) – Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2018.

FAILLA, Zoara (org.). **Retratos da Leitura no Brasil 5**. 1ª ed. Rio de Janeiro: Sextante, 2021. 328p.

FETTER, G. L. **Divulgação tecnológica para agricultores familiares: análise de terminologias sob a ótica da Linguística Sistêmico-Funcional**. 2017.535 f. Dissertação (Mestrado) – Programa de Pós-Graduação em Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2017.

FINATTO, Maria José Bocorny; TCACENCO, Lucas Meireles. Tradução intralingüística, estratégias de equivalência e acessibilidade textual e terminológica. **Tradterm**: TradTerm, São Paulo, v. 37, n. 1, p. 30-62, jan. 2021. Disponível em: www.revistas.usp.br/tradterm. Acesso em: 13 fev. 2021. DOI: <https://doi.org/10.11606/issn.2317-9511.v37i0p30-63>

FINATTO, Maria José Bocorny; MOTTA, Ester. Terminologia e Acessibilidade: novas demandas e frentes de pesquisa. **GTlex**, Uberlândia, p. 316-356, 2019. DOI: <https://doi.org/10.14393/Lex4-v2n2a2017-6>.

FINATTO, Maria José Bocorny. Trabalho com pequenas e grandes amostras textuais: levantamento de terminologias na área Pneumopatias Ocupacionais. In: ISQUERDO, A. N.; DAL CORNO, G. O. M. **A Ciências do Léxico, Lexicologia, Lexicografia, Terminologia**. Campo Grande: Editora da UFMS, 2018, v. 8, p. 347-372.

FINATTO, Maria José Bocorny. **Projeto PorPopular**. Disponível em: http://www.ufrgs.br/textecc/porlexbras/porpopular/download_do_corpus.php. Acesso em: 20 fev. 2017.

FINATTO, M. J. B.; CREMONESE, L.E.; AZEREDO, S. O vocabulário na redação de vestibular: do enfoque estatístico às especificidades da enunciação. In: ABREU, S. (org.) **A redação no vestibular: do leitor ao produtor de texto**. COPERSE/UFRGS. Porto Alegre: Editora da UFRGS, 2008, p.95-108. ISBN978-85-386-0026-8

FINATTO, Maria José Bocorny; EVERS, Aline. Linguística de *Corpus*, Léxico-Estatística Textual e Processamento de Linguagem Natural: perspectiva para estudos de vocabulário em produções textuais. **Revista Gtlex**, [s.l.], v. 1, n. 2, p.271-295, 5 set. 2016. EDUFU - Editora da Universidade Federal de Uberlândia. DOI: <http://dx.doi.org/10.14393/lex2-v1n2a2016-3>.

FINATTO, Maria José Bocorny; EVERS, Aline; STEFANI, Mônica. Letramento científico e simplificação textual: o papel do tradutor no acesso ao conhecimento científico. **Revista Letras**, Santa Maria, v. 26, n. 52, p.135-158, jan./jun. 2016.

FINATTO, Maria José Bocorny; LOPES, Lucelene; CIULIA, Alena. Processamento de Linguagem Natural, Linguística de *Corpus* e Estudos Linguísticos: uma parceria bem-sucedida. **Domínios da Linguagem**, Uberlândia, v. 9, n. 5, p.41-59, dez. 2015. Disponível em: <http://www.seer.ufu.br/index.php/dominiosdelinguagem>. Acesso em: 10 mar. 2017.

FINATTO, Maria José Bocorny e KILIAN, Cristiane Krause. Conceitos básicos da Linguística de Linguagens Especializadas – texto comentário 2. In: Finatto, Maria José B. (Org.); ZILIO, Leonardo (Org.). **Textos e Termos por Lothar Hoffmann, um convite para o estudo das linguagens técnico-científicas**. 1. ed. Porto Alegre: Pallotti, 2015. v. 01. 256 p.

FINATTO, M. J. B.; EVERS, A.; PASQUALINI, B. Desenhos do vocabulário no Português popular escrito: pesquisa, ensino, produtos lexicográficos e terminológicos. In: FINATTO Maria José Bocorny; SCARTON, Carolina Evaristo; ROCHA, Amanda; ALUISIO, Sandra Maria. Características do jornalismo popular: avaliação da inteligibilidade e auxílio à descrição do gênero. In: VIII Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana, 2011, Cuiabá - MT. **Anais do STIL 2011**. Cuiabá : Sociedade Brasileira de Computação, 2011. v. 01. p. 30-39.

FINATTO, Maria José Bocorny. Complexidade textual em artigos científicos: contribuições para o estudo do texto científico em português. **Organon** (UFRGS), v. 50, p. 30-45, 2011.

FINATTO, Maria José Bocorny. Termos, textos e textos com termos: novos enfoques dos estudos terminológicos de perspectiva linguística. In: ISQUERDO, A.N.; KRIEGER, M.G. (Orgs.). **As ciências do léxico: lexicologia, lexicografia, terminologia**, volume II. Campo Grande: Ed. UFMS, p. 341-358, 2004.

FINATTO, M. J. B.; VALE, O. A.; LAPORTE, Éric, “Reconhecimento do vocabulário de jornais populares brasileiros por um dicionário computacional de acesso livre”, *Alfa*, vol. 63, nº 1, abr. 2019. Disponível em: <https://periodicos.fclar.unesp.br/alfa/article/view/11234/8182>

FISCHER, Heloisa. **Clareza em textos de e-gov, uma questão de cidadania**. Rio de Janeiro: Com Clareza, 2018. Livro eletrônico. n.p.

FISCHER, Heloisa. **Impactos da linguagem simples na compreensibilidade da informação em governo eletrônico: o caso de um benefício do INSS – 2021**. 263 f. Dissertação (Mestrado em Design) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Artes e Design, Rio de Janeiro, 2021.

FLESCH, Rudolf. **The art of readable writing**. Nova York: Harper Row, 1949.

FROMM, Guilherme. A construção e análise de *corpora* para a alimentação de um banco de dados: um exemplo. **Domínios de Lingu@gem**, Uberlândia, ano 2, n. 1, p. 1-22, 2008.

FROMM, Guilherme; GRAMA, Daniela Faria; BEILKE, Neubiana Silva Veloso; SANTOS, Candice Guarato. Wordsmith Tools e Sketch Engine: um estudo analítico-comparativo para pesquisas científicas com uso de *corpora* / wordsmith tools and sketch engine. **Revista de Estudos da Linguagem**, [S.L.], v. 28, n. 3, p. 1191-1248, 27 maio 2020. Faculdade de Letras da UFMG. <http://dx.doi.org/10.17851/2237-2083.28.3.1191-1248>

FULGÊNCIO, Lúcia; LIBERATO, Yara. **É possível facilitar a leitura: um guia prático para escrever claro**. São Paulo: Contexto, 2010.

FUNDAÇÃO GETÚLIO VARGAS (FGV). **Estudo da Imagem do Judiciário Brasileiro**. Rio de Janeiro: FGV, 2019. 169 p. Disponível em: https://www.amb.com.br/wp-content/uploads/2020/04/ESTUDO_DA_IMAGEM_DO_JUDICIARIO_BRASILEIRO_CO_MPLETO.pdf. Acesso em: 20 abr. 2020.

GAMBIER, Y. Travail et vocabulaire spécialisés: prolégomènes à une socio-terminologie. **Meta**, vol. 36, n. 1, 1991, p. 8-15.

GAUDIN, François. **Pour une socioterminologie. Des problèmes sémantiques aux pratiques institutionnelles**. Rouen: Publications de l'Université de Rouen, 1993.

GOIÁS. CASA CIVIL. **'Tá na Lei'**: aplicativo do Governo de Goiás ajuda cidadão a conhecer as leis e a defender seus direitos. 2020. Disponível em: <https://www.casacivil.go.gov.br/noticias/9153-t%C3%A1-na-lei-aplicativo-do-governo-de-goias>. Acesso em: 12 fev. 2021.

GRUPO DE LINGUÍSTICA FORENSE (Santa Catarina). **PGI News**. 2016. Disponível em: <http://ppgi.posgrad.ufsc.br/files/2016/03/PPGI-News-March-2016.pdf>. Acesso em: 10 fev. 2017.

HALLIDAY, M. A. K. **Language as Social Semiotic: the social interpretation of language and meaning**. London: Edward Arnold. 1978.

HOEY, M. From concordance to text structure: New uses for computer *corpora*. In: B. LEWANDOSWKA-TOMASZCZYK & P. J. MELIA (org.). **PALC'97 – Practical Applications in Language Corpora**. Lodz: Lodz University Press. 1997.

HOFFMANN, L. Conceitos básicos da Linguística de Linguagens Especializadas. In: Finatto, Maria José B. (Org.); ZILIO, Leonardo (Org.). **Textos e Termos por Lothar Hoffmann, um convite para o estudo das linguagens técnico-científicas**. 1. ed. Porto Alegre: Pallotti, 2015a. v. 01. 256 p.

HOFFMANN, L. Do texto especializado ao gênero textual especializado. In: Finatto, Maria José B. (Org.); ZILIO, Leonardo (Org.). **Textos e Termos por Lothar Hoffmann, um convite para o estudo das linguagens técnico-científicas**. 1. ed. Porto Alegre: Pallotti, 2015b. v. 01. 256 p.

HOFFMANN, L. Características sintáticas e morfológicas de linguagens especializadas. In: Finatto, Maria José B. (Org.); ZILIO, Leonardo (Org.). **Textos e Termos por Lothar Hoffmann, um convite para o estudo das linguagens técnico-científicas**. 1. ed. Porto Alegre: Pallotti, 2015c. v. 01. 256 p.

HOUAISS, A; Villar, M. de S. Dicionário Houaiss Eletrônico. Versão 1.0. **CD-ROM**. Rio de Janeiro: Editora Objetiva, 2007.

INAF/ INSTITUTO PAULO MONTENEGRO. **INAF BRASIL 2018 – Resultados Preliminares**. Disponível em: <https://acaoeducativa.org.br/publicacoes/indicador-de-alfabetismo-funcional-inaf-brasil-2018/>. Acesso em: 15 jun. 2019.

JAKOBSON, R. **On linguistic aspects of translation**. In: Lawrence Venuti ed. The Translation Studies Reader (2012). London: Routledge, 1959.

JANCZURA, Gerson Américo; CASTILHO, Goiara Mendonça de; ROCHA, Nelson Oliveira; VAN ERVEN, Terezinha de Jesus Cordeiro; HUANG, Tin Po. Normas de Concretude para 909 Palavras da Língua Portuguesa. **Psicologia: Teoria e Pesquisa**, Brasília, v. 23, n. 2, p. 195-204, abr.-jun. 2007.

JUST, M. A.; CARPENTER, P. A. **A Capacity Theory of Comprehension: Individual Differences in Working Memory** Psychological Review, 1992, v.. 99, n. 1, 122-149. Disponível em: http://www.ccbi.cmu.edu/reprints/Just_PsychRev-1992_capacity-theory.pdf. Acesso em: 13 fev. 2021.

KLEIMAN, A. **Texto e leitor: aspectos cognitivos da leitura**. Campinas/SP: Pontes, 1997.

KLEIMAN, A. **Oficina de leitura: teoria e prática**. Campinas/SP: Pontes, 2002.

KRIEGER, M. G.; FINATTO, M. J. B. **Introdução à terminologia: teoria e prática**. São Paulo: Contexto, 2004.

LAGES, Margarida. Os desafios da linguagem jurídica para uma comunicação eficiente. **Revista do Tribunal Regional do Trabalho da 3. Região**, Belo Horizonte, v. 55, n. 85, p. 169-208, jan./jun. 2012.

LEAL, S. E. **Predição da complexidade sentencial do português brasileiro escrito, usando métricas linguísticas, psicolinguísticas, e de rastreamento ocular**. 2019. 132 p. Monografia (Doutorado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2019.

LEECH, G. *Corpora* and theories of linguistic performance. In: J. SVARTVIK (org.). **Directions in Corpus Linguistics**. Proceedings of Nobel Symposium 82, Stockholm, 4-8 August 1991. Berlin, New York: De Gruyter. 1992.

LEFFA, V. J. **Fatores da compreensão na leitura**. Projeto ELO, Ensino de línguas online: 1996. Disponível em: www.leffa.pro.br. Acesso em: 20 fev. 2017.

LOPES, L.; VIEIRA, R. Estimating term domain relevance through term frequency, disjoint *corpora* frequency - tf-dcf. **Knowledge-based Systems**, p. 237-249, 2016.

LOPES, Lucelene. **Extração automática de conceitos a partir de textos em língua portuguesa**. 2012. 156 f. Tese (Doutorado em Ciência da Computação) -PUCRS – Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2012.

MACIEL, Roger Luiz. **Linguagem jurídica: é difícil escrever direito?** Disponível em: www.conjur.com.br/. Acesso em: 30 nov. 2009.

MACIEL, Anna Maria Becker. **Para o reconhecimento da especificidade do termo jurídico**. 2001. 291 f. Tese (Doutorado), Programa de Pós-Graduação em Letras, Universidade Federal do Rio Grande do Sul, 2001.

MAGALHÃES PINTO, Oriana Piske de Azevedo. 2005. Pela simplificação da linguagem jurídica: ninguém valoriza o que não entende. Disponível em: http://www.amb.com.br/index.asp?secao=artigo_detalhe&art_id=220 . Acesso em: 18 set. 2019.

MAIOR, Jorge Luiz Souto. Opinião pública e direito do trabalho: tentando transpor as barreiras da comunicação. **Revista Iob**, São Paulo, v. 19, n. 221, p.32-46, nov. 2007.

MCENERY, T. & A. WILSON. **Corpus Linguistics**. Edinburgh: Edinburgh University Press. 1996.

MESQUITA, Nely Silva da Motta. **Legibilidade Textual nos Cadernos Didáticos de EJA: questões de processamento da leitura**. 2011. 97 f. Dissertação (Mestrado em Letras). Departamento de Letras, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2011.

MONTORO, André Franco. (1998) **Curso de Teoria Geral do Direito**. aula 2; A linguagem do direito: semântica, sintática e pragmática. PUCSP- Faculdade Paulista de Direito, Pós-Graduação em Direito. Disponível em: <http://www.dialdata.com.br/ilam/aula2>. Acesso em: 13 fev. 2021.

MONZÓN, Andrea Jessica Borges. **Terminologia do inglês da Ciência da Computação e seus desdobramentos em cursos técnicos e tecnológicos de Informática nos Institutos Federais**. 2017. 270 f. Tese (Doutorado) – Curso de Programa de Programação em Letras, Instituto de Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2017.

MOTTA, Ester. Sentenças Judiciais e Acessibilidade Textual e Terminológica. **Domínios de Lingu@gem**, v. 15, n. 3, p. 761-813, 13 ago. 2021. DOI: <https://doi.org/10.14393/dl>

MOTTA, Ester. 2018. Índices de complexidade textual em Sentenças dos Juizados Especiais Cíveis do Poder Judiciário do Estado do Rio Grande do Sul. **Revista Inventário**, Salvador - BA, n. 21, p.19-34, julho 2018.

MOTTA-ROTH. Popularização da ciência como prática social e discursiva. In: MOTTA-ROTH, D.; GIERING, M. E. (Orgs.). **Discursos de popularização da ciência**. Santa Maria, RS: PPGL Editores, 2009. 343p. (Coleção HiperS@beres, 1). ISSN 2177-6385.

MOZDZENSKI, Leonardo P. Análise crítica do discurso jurídico: uma proposta de investigação. **Revista da Faculdade de Direito de Caruarú**, v. 42, p. 1-20, 2010.

MOZDZENSKI, Leonardo P. Desconstruindo a linguagem jurídica: multimodalidade e argumentatividade visual nas cartilhas de orientação legal. **Veredas** (UFJF, Impresso), v. 8, p. 91-106, 2004.

NETO, Liberato Bonadia. **Juizados Especiais Cíveis – evolução – competência e aplicabilidade – algumas considerações**. Disponível em: <http://www.jornaljurid.com.br/doutrina/civil/juizados-especiais-civeis-evolucao-competencia-e-aplicabilidade-algumas-consideracoes>. Acesso em: 02 mar. 2017.

NORTHFLEET, Ellen Gracie. [**Íntegra do discurso de posse da ministra Ellen Gracie como presidente do STF**]. Disponível em: <http://www.stf.jus.br/portal/cms/verNoticiaDetalhe.asp?idConteudo=100112>. Acesso em: 10 fev. 2016

NOVODVORSKI, A.; BOCORNY FINATTO, M. J. Linguística de *Corpus* no Brasil: uma aventura mais do que adequada. **Letras & Letras**, [S. l.], v. 30, n. 2, p. 7–16, 2014. DOI: 10.14393/LL60-v30n2a2014-1. Disponível em: <https://seer.ufu.br/index.php/letraseletras/article/view/28516>. Acesso em: 6 maio. 2022.

OLIVEIRA, Márcia Regina Alves Ribeiro. Jornal Popular X Jornal Tradicional: Análise léxico-gramatical da notícia a partir da Linguística de *Corpus* Um estudo de casos dos jornais cariocas “O Globo” e “O Dia”. **Veredas On-line – Linguística de Corpus e Computacional**, Juiz de Fora, v. 13, n. 2, p.7-19, fev. 2009. Disponível em: <http://www.ufjf.br/revistaveredas/files/2009/11/ARTIGO-Márcia-Regina-Alves-Ribeiro-Oliveira1.pdf>. Acesso em: 02 abr. 2017.

PASQUALINI, B. F. **CORPOP: um corpus de referência do português popular escrito do Brasil**. 2018. 250 f. Tese (Doutorado) - Curso de Letras, Programa de Pós-graduação em Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2018.

PASQUALINI, Bianca Franco. **Leitura, Tradução e Medidas de Complexidade Textual em Contos da Literatura para Leitores com Letramento Básico**. 2012. 155f. Dissertação. (Mestrado) – Programa de Pós-Graduação em Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2012.

PASQUALINI, Bianca Franco. Análise de Índices de complexidade textual em dois textos de pediatria e suas traduções para o inglês: um estudo exploratório. **Palimpsesto**, Rio de Janeiro, n. 13, ano 10, 2011. Disponível em: <<http://www.pgletras.uerj.br/palimpsesto/num13/estudos/estudos.htm>>. Acesso em: 30 mar. 2017.

PARAGASSU, Liana. **Tradução Especializada Acessível (TEA): revisão do tema e proposta de disciplina para cursos de graduação em tradução**. 2018. 300 f. Dissertação (Mestrado) - Curso de Letras, Programa de Pós-graduação em Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2018.

PAUMIER, S. **Unitex 1.2: Manual do Usuário**. Traduzido por Amanda Lopes Pietrobom. Paris: Université Paris-Est Marne-la-Vallée, 2007.

PEREIRA, V. W.; BARETTA, D. Compreensão literal e inferencial em alunos do Ensino Fundamental. **Signo**, Santa Cruz – RS, v. 43, p. 53/77-61, 2018.

PERFETTI. The acquisition of Reading comprehension skill. SNOWLING, Margaret J.; HULME, Charles (Eds.) In: **The science of Reading: a handbook**. Oxford: Blackwell, 2005.

PERINI, M. A. **Sintaxe Portuguesa: metodologia e funções**. São Paulo: Ática. 1994.

PERINI, M. A. Tópicos discursivos e legibilidade. In: PERINI, M. A. (coord.) **Definição linguística da legibilidade**. Belo Horizonte: Relatório de Pesquisa ao Inep, 1982.

PERNA, Cristina Lopes (Org.); DELGADO, Heloísa Koch (Org.); FINATTO, Maria José Bocorny (Org.). **Linguagens especializadas em corpora: modos de dizer e interfaces**. 1. ed. Porto Alegre - RS: EDIPUCRS, 2010. v. 01. 255p.

PETRI, Maria José Constantino. **Manual de Linguagem Jurídica**. 3. ed. São Paulo: Saraiva, 2017.

PLAIN LANGUAGE ACTION AND INFORMATION NETWORK (Estados Unidos). **Federal Plain Language Guidelines**. 2011. Disponível em: <<http://www.plainlanguage.gov/howto/guidelines/FederalPLGuidelines>>. Acesso em: 20 out. 2016.

PLAIN LANGUAGE ASSOCIATION INTERNATIONAL (PLAIN) (Canadá). **PLAIN**. Disponível em: <http://plainlanguagenetwork.org/plain-language/what-is-plain-language>. Acesso em: 12 fev. 2017.

RIO GRANDE DO SUL. Poder Judiciário do Rio Grande do Sul. Processo nº 015/3.14.0000039-0, Comarca de Gravataí, julgado em 07-07-2015.

RIO GRANDE DO SUL. Poder Judiciário do Rio Grande do Sul. Processo nº 9000378-32.2015.8.21.0027, Comarca de Santa Maria, julgado em 23-07-2015.

SALGUEIRO PARDO, Thiago Alexandre; VOLPE NUNES, Maria das Graças. On the Development and Evaluation of a Brazilian Portuguese Discourse Parser. **Revista de Informática Teórica e Aplicada**, Porto Alegre, RS, v. 15, n. 2, p. 43-64, dec. 2008. ISSN 21752745. Disponível em: <https://seer.ufrgs.br/rita/article/view/rita_v15_n2_p43-64/4485>. Acesso em: 01 fev. 2021. DOI: <https://doi.org/10.22456/2175-2745.7015>

SANTANA, Samene Batista Pereira. A linguagem jurídica como obstáculo ao acesso à justiça. Uma análise sobre o que é o Direito engajado na dialética social e a consequente desrazão de utilizar a linguagem jurídica como barreira entre a sociedade e o Direito/Justiça. Disponível em: http://www.ambito-juridico.com.br/site/index.php/?n_link=revista_artigos_leitura&artigo_id=12316&revista_caderno=24. Acesso em: 06 fev. 2017.

SANTOS, D. Corporizando algumas questões. In: TAGNIN, S.; VALE, O. A. **Avanços da Linguística de Corpus no Brasil**. São Paulo: Humanitas, 2008. p. 41-66.

SANTOS, Jusinete. **Avaliação das habilidades de leitura no nível superior a partir da análise de provas de nivelamento**. 2013. 166 f. Dissertação (Mestrado) – Programa de Pós-Graduação em Estudos da Linguagem, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2013.

SANTOS, Leandro Borges dos; DURAN, Magali Sanches; HARTMANN, Nathan Siegle; CANDIDO JUNIOR, Arnaldo; PAETZOLD, Gustavo Henrique; ALUÍSIO, Sandra Maria. A lightweight regression method to infer psycholinguistic properties for Brazilian Portuguese. In: International Conference on Text, Speech, and Dialogue - TSD 2017, Proceedings, volume 10415 of Lecture Notes in Artificial Intelligence, pages 281–28. Springer.

SASSAKI, Romeu Kazumi. **Inclusão: construindo uma sociedade para todos**. Rio de Janeiro: WVA, 1997. 176 p.

SCARTON, Carolina Evaristo; ALUÍSIO, Sandra Maria. Análise da Inteligibilidade de textos via ferramentas de Processamento de Língua Natural: adaptando as métricas do Coh-Metrix para o Português. **Linguamática** (Revista para o Processamento Automático das Línguas Ibéricas), v. 2, n. 1, p. 45-61, 2010. Disponível em: <http://linguamatica.com/index.php/linguamatica/article/viewfile/44/59>. Acesso em: 20 fev. 2017.

SILVA, De Plácido e. **Vocabulário Jurídico**. 22.ed. Rio de Janeiro: Forense, 2007. (eletrônico)

SILVA, Asafe Davi Cortina. **Textos de Divulgação para Leigos sobre o Transtorno do Estresse Pós-Traumático em Português: alternativas para a acessibilidade textual e terminológica**. 2018. 427 f. Dissertação (Mestrado) - Curso de Letras, Programa de Pós-graduação em Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2018.

SILVA, Bruna Rodrigues da. **Vocabulário escrito de estudantes de escolas públicas do Rio Grande do Sul: um estudo léxico-estatístico**. 2021. 321 f. Dissertação (Mestrado) - Curso de Letras, Instituto de Letras - Programa de Pós-Graduação em Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2021.

SILVA, R. S. **Diagramação: o planejamento visual gráfico na comunicação impressa**. São Paulo: Summus Editorial, 1985.

SINCLAIR, J. 2005. *Corpus and Text - Basic Principles*. In: M. WYNNE (ed.), **Developing Linguistic Corpora: a Guide to Good Practice**. Oxford, Oxbow Books, p. 1-16. Disponível em: <http://ahds.ac.uk/linguistic-corpora/>. Acesso em: 30 out. 2006.

SOARES, Magda. Letramento e alfabetização: as muitas facetas. **Revista Brasileira de Educação**. Jan/Fev/Mar/Abr 2004, n. 25, p.5-17. Disponível em <http://www.scielo.br/pdf/rbedu/n25/n25a01.pdf>. Acesso em: 16 ago. 2018.

SOUTO MAIOR, Jorge Luiz. Opinião pública e direito do trabalho: tentando transpor as barreiras da comunicação. **Revista IOB**, Ano XIX, n. 221, nov./2007.

SOUZA, Renata Martins. A elitização da linguagem jurídica como obstáculo ao acesso à Justiça. **CONJUR**, 2020. Disponível em: <https://www.conjur.com.br/2020-set-29/tribuna->

[defensoria-elitizacao-linguagem-juridica-obstaculo-acesso-justica](#) . Acesso em: 07 maio 2022

TESHEINER, André Luís de Aguiar. Linguagem Simples e Visual Law. In: COELHO, Alexandre Zavaglia; SOUZA, Bernardo de Azevedo e (org.). **Legal Design e Visual Law no Poder Público**. São Paulo: Thomson Reuters - Revista dos Tribunais, 2021. Cap. 4. p. 113-127.

VERÍSSIMO, L. F. Quem. **Zero Hora**. Porto Alegre, 9 abr. 2018.

VIVIAN, Ester M. **A Variação no léxico jurídico: um estudo aplicado de unidades terminológicas complexas**. 2010. 135f. Dissertação (Mestrado) Programa de Pós-Graduação em Letras, UFRGS, Porto Alegre, 2010.

WOLF, Maryanne. **Cérebro no mundo digital: os desafios da leitura na nossa era**. Edição do Kindle.

WÜSTER, E. Introducción a la teoría general de la terminología y a la lexicografía terminológica. Barcelona: Universitat Pompeu Fabra, 1998.

XAVIER, Ronaldo Caldeira Xavier. **Português no Direito: linguagem forense**. Rio de Janeiro: Forense, 2002. p. 1.

ZARISKI, Archie. **What is legal literacy?: Examining the concept and objectives of legal literacy**. Athabasca: Athabasca University, 2011. Disponível em: http://lgst.athabascau.ca/courses/documents/LGST249_sample.pdf. Acesso em: 21 nov. 2018.

ZETHSEN, Karen Korning. Intralingual translation: an attempt at description. **Meta** 54(4). pp. 795-812, 2009. DOI: 10.7202/038904ar.

ZILIO, L.; WILKENS, R.; FAIRON, C. **PassPort: a dependency parsing model for portuguese**. In: PROPOR - International Conference of Computational Processing of the Portuguese Language, 13, 2018, Canela. *Proceedings [...]*. Cham: Springer, 2018. p. 479-489. Disponível em: https://link.springer.com/chapter/10.1007/978-3-319-99722-3_48. Acesso em: 15 abr. 2020.

ZILIO, Leonardo. Terminologia textual e linguística de *corpus*: um estudo em parceria. In: PERNA, Cristina Lopes; DELGADO, Heloísa Koch; FINATTO, Maria José Bocorny. **Linguagens especializadas em corpora: modos de dizer e interface de pesquisa**. Porto Alegre: Edipucrs, 2010. p. 184-201. Disponível em: <http://www.pucrs.br/edipucrs>. Acesso em: 15 fev. 2017.

ZILIO, Leonardo. **VERBLEXPOR: um recurso léxico com anotação de papéis semânticos para o português**. 2015. 197 f. Tese (Doutorado) - Curso de Letras, Programa de Pós-graduação em Letras, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2015.

ANEXOS**ANEXO 1 - TEXTO ORIGINAL DA SENTENÇA SIMPLIFICADA NO CAPÍTULO 10****PROPOSTA DE SENTENÇA****Comarca de RIO GRANDE****Juizado Especial Cível**

Processo nº: xxxxxxxxxxxxxxxx**Natureza:** Contratos de Consumo :: Consórcio**Autor:** xxxxxxxxxxxxxxxx**Ré:** xxxxxxxx Administradora de Consórcios Nacional Ltda.**Prolatora:** xxxxxxxxxxxxxxxx Juíza Leiga**Data:** 12/02/2016**Vistos , etc.**

Em que pese a dispensa do relatório consoante art. 38, in fine, da Lei nº 9.099/95, faz-se um breve arrazoado dos fatos.

Trata-se ação indenizatória promovida por XXXXX em face de XXX Administradora de Consórcios Nacional Ltda., em que narrou ter ingressado em um consórcio, cujo o bem ofertado é um automóvel, em cem parcelas de R\$ 397,61 (trezentos e noventa reais e sessenta e um centavos) acrescidos de tarifas mensais. Referiu, contudo, já ter efetuado o pagamento de quinze parcelas e que continua sendo cobrado o valor referente a Taxa de Administração Mensal, que conforme o contrato seria cobrado da primeira à quinta parcela. Aduz ter entrado em contato com a requerida para esclarecer tal fato e solicitar a retirada dessa taxa de seus boletos futuros, porém não obteve sucesso. Requereu a condenação da ré a restituição em dobro do valor pago e a indenização a título de danos morais.

No mérito, a análise do cotejo probatório carreado aos autos conduz à improcedência por parte autora do pedido articulado na peça preambular. Fundamento.

Inicialmente, calha salientar que a relação que envolve os litigantes é eminentemente de consumo, assumindo cada qual a posição de consumidor como destinatário final e prestador de serviços. Corolário lógico é a aplicação das regras previstas no CDC, inclusive a inversão do ônus *probandi*.

Consequentemente, no presente caso cabe inverter-se o ônus da prova conforme previsão do art. 6º, VIII do CDC, recaindo assim o ônus da comprovação sobre a demandada.

Da análise das teses levantadas pelas partes e documentos juntados, verifica-se que a requerida agiu de boa-fé contratual, exibindo um compromisso de fidelidade e cooperação nas relações contratuais. Também é visível que a requerida cumpriu com o dever da informação, conforme documento em anexo (fls. 48).

Na opinião de Rizzato Nunes, entendemos o princípio da informação preceituado no Código de Proteção e Defesa do Consumidor como:

“Dever de informar: com efeito, na sistemática implantada pelo CDC, o fornecedor está obrigado a prestar todas as informações acerca do produto e do serviço, suas características, qualidades, riscos, preços e etc., de maneira clara e precisa, não se admitindo falhas ou omissões.”

Com efeito, constava no contrato assinado pelo autor que o valor total da taxa de administração é de R\$ 16.500,00 (dezesesseis mil e quinhentos reais), equivalendo a 16,5% do montante, e que seria pago nas cinquenta primeiras parcelas. Assim, se o autor pagasse apenas até a 5ª parcela a taxa de administração, teria pago somente 1,65%, o que não estaria de acordo com o estipulado pelo contrato.

Inegável que a taxa de administração compactuada é de 16,5%, portanto é cobrada nas 50 primeiras parcelas para configurar o valor estipulado de R\$ 16.500,00 (dezesesseis mil e quinhentos reais).

Todavia, houve um erro de preenchimento no contrato onde ficou escrito que a taxa de administração mensal seria da parcela 01 a 05, gerando um erro escusável, que é um erro justificável ao se levar em conta as circunstâncias do fato.

No caso em apreço, é possível verificar no preenchimento do contrato que não haveria a taxa de administração mensal a partir da parcela 50 até a 100, o que torna o primeiro erro escusável mediante raciocínio lógico.

Com relação à condenação por danos morais, tem-se que este não comporta acolhimento. Com efeito, no caso sub judice, tem-se que não restaram configurados na espécie. Em que pese não se reconheça qualquer elemento que configure situação de abalo moral e psicológico a ensejar a indenização pretendida, já que não restou configurado o descumprimento contratual.

A situação exposta pelo autor, quando muito, caracteriza incômodo, aborrecimento ou desconforto, em maior ou menor grau conforme a suscetibilidade subjetiva de cada um, e nenhuma das circunstâncias do caso, consideradas em si mesmas ou reunidas umas com as outras, chegam a configurar situação característica de dano moral. Há contratempus e transtornos que são próprios da vida em sociedade.

Nesse sentido:

APELAÇÃO CÍVEL. ALIENAÇÃO FIDUCIÁRIA. CONSORCIO. DIREITO DO CONSUMIDOR/CONTRATO DE CONSUMO. CASO CONCRETO. AÇÃO DE REPETIÇÃO DE INDÉBITO CUMULADA COM PLEITO DE DANOS MORAIS. ADESÃO AO PLANO DE CONSORCIO EM ANDAMENTO. CONTEMPLAÇÃO POR SORTEIO. CARTA DE CRÉDITO. ADIMPLIMENTO DAS PARCELAS EM ABERTO. PREVISÃO CONTRATUAL E REGRAMENTO EXPRESSAMENTE PACTUADO, COM CIÊNCIA DO

CONSORCIADO. ERRO ESCUSÁVEL NA EMISSÃO DA CARTA DE CRÉDITO NO VALOR INTEGRAL. PRETENSÃO DO AUTOR DESACOLHIDA ANTE O PRIMADO DA BOA-FÉ CONTRATUAL. SENTENÇA DE IMPROCEDÊNCIA MANTIDA EM SEUS TERMOS NESTE GRAU

RECURSAL. RECURSO DESPROVIDO. (Apelação Cível Nº 70058148784, Décima Terceira Câmara Cível, Tribunal de Justiça do RS, Relator: Angela Terezinha de Oliveira Brito, Julgado em 03/04/2014)

Assim, feitas essas breves considerações, entende-se não merecer prosperar o pedido de condenação em danos morais requerido pelo autor, tendo em conta se tratar de mero dissabor a situação experimentada.

ANTE O EXPOSTO, com base na fundamentação retro e fulcro no art. 269, I, do CPC, opina-se pela IMPROCEDÊNCIA do pedido articulado pela autora, no sentido de julgar improcedente o pedido de reparação a título de danos morais. Sem sucumbência ou condenação de honorários.

Remeta-se à apreciação da Presidência. Intimem-se.

Rio Grande, 12 de fevereiro de 2016.

Juíza Leiga.

Local e Data: Rio Grande, 16 de dezembro de 2015.

SENTENÇA HOMOLOGATÓRIA

Nos termos do art. 40 da Lei nº 9.099/95, homologo a proposta de decisão, para que produza efeitos como sentença. Sem custas e honorários, na forma da Lei. As partes consideram-se intimadas a partir da publicação da decisão, caso tenha ocorrido no prazo assinado; do contrário, a intimação terá de ser formal.

Rio Grande, 16 de dezembro de 2015.

Juiz de Direito.

ANEXO 2 - TEXTO SIMPLIFICADO DO EXEMPLAR DA SENTENÇA DO CAPÍTULO 10

PROPOSTA DE SENTENÇA

Comarca XXXXXXXX

Juizado Especial Cível

Vistos , etc.

Mesmo que o art. 38 da Lei 9.099/95 não exija o relatório, faremos um pequeno resumo dos fatos.

XXXXXXXX ingressou com uma ação de indenização contra XXXXXXXXX, porque entrou num consórcio de um automóvel em cem parcelas de R\$ 397,61 (trezentos e noventa reais e sessenta e um centavos) mais tarifas mensais.

Referiu, [...]. Conta que entrou em contato a requerida para esclarecer tal fato e solicitar a retirada dessa taxa de seus boletos futuros, porém não obteve sucesso. Requereu a condenação da ré à restituição em dobro do valor pago e à indenização a título de danos morais.

[...]

No mérito, a análise das provas trazidas aos autos não permite que o pedido da petição inicial seja atendido.

Fundamento.

Inicialmente, salienta-se que a relação entre as partes deste processo é de consumo: uma das partes é o consumidor, e a outra é o prestador de serviços. Portanto, as regras do Código de Defesa do Consumidor (CDC), inclusive a inversão do ônus da prova, devem ser aplicadas no caso. A inversão do ônus da prova significa que quem tem que provar é a parte mais forte na relação, e não obrigatoriamente quem alega. Por isso, seguindo o que diz o inc. VIII do art. 6º do CDC, quem deverá provar neste processo é a empresa responsável pelo consórcio.

Da análise das razões trazidas pelas partes e documentos juntados, verifica-se que a empresa responsável pelo consórcio agiu de boa-fé, exibindo um compromisso de fidelidade e cooperação nas relações contratuais. Também cumpriu com o dever da informação, conforme documento em anexo (fl. 48).

Na opinião de Rizzato Nunes, o princípio da informação estabelecido no CDC é entendido como [...].

No caso, é possível verificar no contrato que não haveria a taxa mensal a partir da parcela 50 até a 100. Assim, o primeiro erro apontado não se deu de forma proposital, conforme se verifica da leitura do texto.

Quanto ao pedido de condenação por danos morais, ele não pode ser atendido, porque tais danos não foram comprovados no caso. Não existe qualquer elemento que represente situação de abalo moral e psicológico para se exigir a indenização solicitada. O contrato foi cumprido.

[...].

[...].

Assim, feitas essas breves considerações, entendemos que o pedido de condenação em danos morais solicitado pelo autor não pode ser atendido. A situação ocorrida aqui representou mero aborrecimento.

ANTE O EXPOSTO, com base na fundamentação acima e com base no art. 269, I, do CPC, negamos o pedido de reparação a título de danos morais.

Sem sucumbência ou condenação de honorários. Ou seja, quem perdeu não precisa pagar a quem venceu e também não é preciso pagar nada para o advogado.

Envie-se à apreciação da Presidência.

Intimem-se.

SENTENÇA HOMOLOGATÓRIA

Nos termos do art. 40 da Lei nº 9.099/95, homologo a proposta de decisão, para que produza efeitos como sentença. Sem custas e honorários, na forma da Lei. As partes consideram-se intimadas a partir da publicação da decisão, caso tenha ocorrido no prazo assinado; do contrário, a intimação terá de ser formal.

XXXXXXXX, 16 de dezembro de 2015.

Juiz de Direito.

ANEXO 3 – CORPUS DE ESTUDO

Como os arquivos com o CE e com as listas de palavras utilizadas em nossa pesquisa ficaram muito extensos, disponibilizamos um *link* de acesso para cada um deles. A seguir, colocamos cada um desses *links*.

O link para o arquivo com as 440 Sentenças analisadas nesta pesquisa é o que segue:

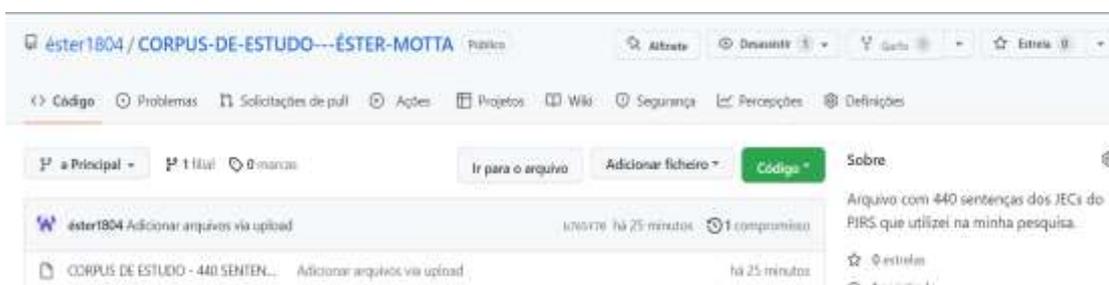
<https://github.com/ester1804/CORPUS-DE-ESTUDO---ESTER-MOTTA.git>

Faremos aqui um passo a passo para o acesso a este e aos demais arquivos.

Passo 1: Clique no *link* disponibilizado (<https://github.com/ester1804/CORPUS-DE-ESTUDO---ESTER-MOTTA.git>).

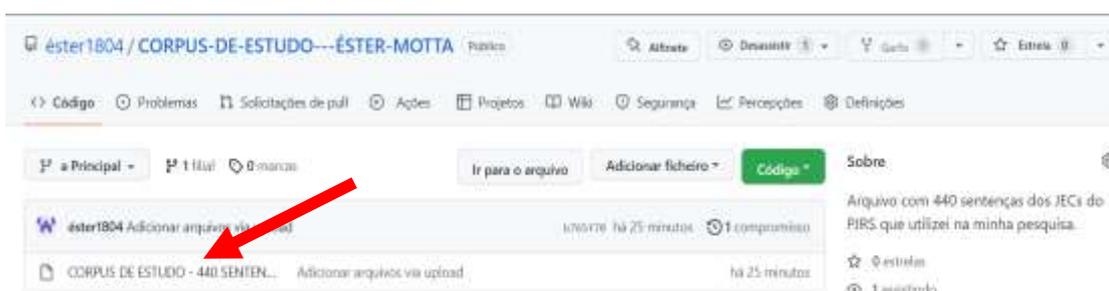
A seguinte tela surgirá:

Figura A.1– Passo 1



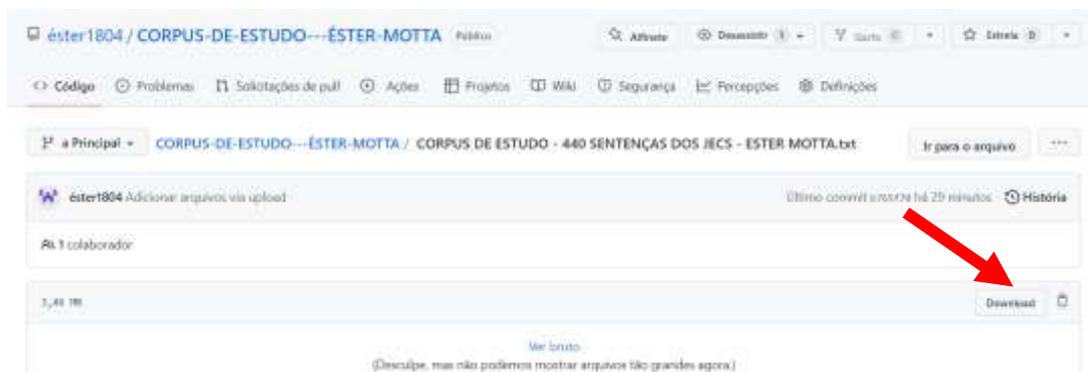
Passo 2: Clique onde a seta indica na imagem abaixo.

Figura A.2 – Passo 2



Passo 3: Outra tela abrirá. Ela poderá colocar o arquivo visível ou disponível para download. Para baixar, é só clicar onde a seta indica na figura abaixo.

Figura A.3 – Passo 3



ANEXO 4 - LISTAS DE PALAVRAS UTILIZADAS NA PESQUISA

CORPUS DE ESTUDO

O *link* para a lista de palavras do CE é: <https://github.com/ester1804/LISTA-PAL-CORPUS-DE-ESTUDO.git>

CORPUS DE CONTRASTE DO JORNAL DIÁRIO GAÚCHO

O *link* para a lista de palavras do DG é: <https://github.com/ester1804/LISTA-PAL-JORNAL-DGA-CHO.git>

CORPUS DE CONTRASTE DO JORNAL MASSA

O *link* para a lista de palavras do JM é: <https://github.com/ester1804/LISTA-PAL-JORNAL-MASSA.git>

CORPUS DE CONTRASTE DO JORNAL ZERO HORA

O *link* para a lista de palavras do ZH é: <https://github.com/ester1804/LISTA-PAL-JZEROHORA.git>

CORPUS DE CONTRASTE DO JORNAL FOLHA DE SÃO PAULO

O *link* para a lista de palavras do FSP é: <https://github.com/ester1804/LISTA-PAL-JFOLHAS-OPAULO.git>

CORPUS DE REFERÊNCIA 1 – OBRAS LITERÁRIAS

O *link* para a lista de palavras do CR1 é: <https://github.com/ester1804/LISTA-PAL-OBRASLITER-RIAS.git>

Caso persistam problemas de acesso, encaminhar e-mail para: estermottac@gmail.com.